

DOI: 10.19783/j.cnki.pspc.241555

基于 Linear Transformer-CNN 模型的继电保护系统 缺陷文本深度挖掘技术

戴志辉, 贺欲飞, 张洪嘉, 张富泽, 韩笑

(河北省分布式储能与微网重点实验室(华北电力大学), 河北 保定 071003)

摘要: 针对当前继电保护系统中缺陷文本数据规模庞大, 而传统挖掘方法普遍存在特征提取不充分、语义识别精度有限、运行效率不高等问题, 提出一种基于 Linear Transformer-CNN 模型的缺陷文本深度挖掘方法。首先, 对海量缺陷文本进行预处理, 将处理好的单词输入到 MacBERT 中生成综合词向量。之后, 将线性注意力机制引入 Transformer 中, 以提高整体运行效率。然后, 添加多层 CNN 模块以弥补 Linear Transformer 对缺陷文本特征提取的不足。最后, 将综合词向量输入到多层 CNN 和 Linear Transformer 模块中, 分别提取缺陷文本局部关键特征和长距离语义特征, 经融合后通过 SoftMax 进行缺陷文本分类。实验结果表明, 相比于传统的文本挖掘方法, 该方法的训练和测试时间更短, 且分类准确率达到 94.24%, 实现了对缺陷文本的快速精准分类。

关键词: 继电保护系统; 缺陷文本; 深度学习; 注意力机制; 文本分类

Deep mining technology for relay protection system defect text based on Linear Transformer-CNN model

DAI Zhihui, HE Yufei, ZHANG Hongjia, ZHANG Fuze, HAN Xiao

(Hebei Key Laboratory of Distributed Energy Storage and Microgrid (North China Electric Power University), Baoding 071003, China)

Abstract: In view of the large volume of defective texts in the current relay protection systems and the limitations of traditional mining methods, such as insufficient text feature extraction, inaccurate semantic recognition, and low operation efficiency, a defect text deep mining method based on the Linear Transformer-CNN model is proposed. First, massive defect text data are preprocessed, and the processed words are input into MacBERT to generate comprehensive word embeddings. Next, a linear attention mechanism is introduced into the Transformer to improve overall operation efficiency. Then, a multi-layer CNN module is added to compensate for the Linear Transformer's limited ability to extract defect text features. Finally, the comprehensive word embeddings are fed into the multi-layer CNN and Linear Transformer modules to extract local key features and long-distance semantic features of defect texts, respectively. The fused features are then classified using a SoftMax layer. Experimental results show that, compared with traditional text mining methods, the training and testing time of the proposed method is shorter, and the classification accuracy reaches 94.24%, enabling fast and accurate classification of defect texts.

This work is supported by the National Natural Science Foundation of China (No. 51877084).

Key words: relay protection system; defect text; deep learning; attention mechanism; text classification

0 引言

继电保护系统作为电力系统的第一道防线, 其能够迅速、准确地定位并隔离故障是保障电网安全稳定运行的前提^[1-3]。随着电网规模的不断扩大和整

体复杂程度的提高^[4-6], 继电保护系统的规模急剧增加, 随之而来的缺陷文本数量也呈指数增长。这些看似杂乱无章的文本中蕴藏着丰富的缺陷信息。目前, 继电保护系统缺陷文本分类主要依赖人工^[7-9], 这种传统的分类方式难以满足海量缺陷记录的处理需求, 且分类结果受限于人员的主观判断。引入自然语言处理技术(natural language processing, NLP)^[10-12],

基金项目: 国家自然科学基金项目资助(51877084)

可以实现对缺陷记录的客观、高效分类, 以提升运维管理水平^[13-14]。

深度学习的发展为自然语言处理带来了巨大变化。国外学者 Krizhevsky 等人开创性地将卷积神经网络(convolutional neural network, CNN)应用于图像分类, 随后 Kim 等人将其拓展至文本分类, 提出了 TextCNN^[15]。然而, TextCNN 在处理长文本时存在局限性, 无法充分捕捉上下文信息。为了解决这一问题, 循环神经网络应运而生, 但梯度消失问题限制了其性能。长短时记忆网络通过引入门控机制缓解了这一问题, 但仍无法利用未来的信息^[16]。双向长短期记忆网络通过双向建模进一步提升了模型的表达能力, 但对于长文本的重点信息提取仍有不足。引入注意力机制的 Transformer 模型能有效地解决这一问题^[17], 使得模型能够更加关注全文中的关键信息。然而, 注意力机制的二次计算复杂度严重影响了其运算速度, 若能对其进行改进, 则可大幅提高缺陷文本的分析效率。

国内学者在缺陷文本挖掘领域已取得了一定进展。然而, 现有研究中部分模型并未利用注意力机制来学习特征权重, 从而无法把握长距离语义信息, 另一部分模型则因引入注意力机制导致模型过于复杂, 从而影响分类效果或降低模型效率。文献[18]基于隐马尔可夫法及 K 最近邻算法完成了断路器全寿命健康状态指数的展示, 但隐马尔可夫法需要依靠词库对单词进行编码, 不适用于拥有大量专业名词的电力行业。文献[19]分析了缺陷文本的特点, 并基于此构建了 TextCNN 缺陷文本分类模型, 但 CNN 模型只能提取局部特征信息, 在处理长文本时无法充分捕捉上下文信息, 在一定程度上限制了其性能。文献[20]利用双向长短期记忆网络提取缺陷文本的深层语义特征, 在一定程度上解决了循环神经网络梯度爆炸或消失的问题, 然而其主要关注序列中的局部依赖关系, 对于序列中全局的、长距离的依赖关系捕捉能力仍然有限。文献[21]在双向长短期记忆模型的基础上引入了注意力机制, 在处理中长文本上具有明显优势。文献[22]采用组合模型, 在文献[21]的基础上添加 CNN 并行模块, 旨在进一步提升模型的准确度。然而, 复杂的模型对计算量有着较大的要求, 因此需要对模型进一步优化, 做到效率与准确率的提升。

除此之外, 由谷歌提出的基于 Transformers 的双向编码器表征(bidirectional encoder representations from Transformers, BERT)模型标志着自然语言处理技术的又一次重大革新^[23]。作为 Word2Vec 的继任者, BERT 及其各种改进模型凭借双向编码器结构,

在缺陷文本预处理与词向量生成上表现优异。文献[24]利用 BERT 对文本进行分词处理, 有效提高了电力领域中文文本的分词质量。

针对上述问题, 本文提出了一种基于 Linear Transformer-CNN 模型的继电保护系统缺陷文本深度挖掘方法。首先, 利用 MacBERT 模型实现继电保护系统缺陷文本向量化。之后, 将线性注意力机制引入 Transformer 得到 Linear Transformer 模型, 以降低复杂度。然后, 加入多层 CNN 模型以提高模型局部特征提取能力。最后, 将从 Linear Transformer 和 CNN 得到的特征融合, 并通过 SoftMax 得到分类结果。实例结果证明, 与传统深度学习模型对比, Linear Transformer-CNN 模型训练时间更短, 分类准确度更高, 能更好地胜任继电保护系统缺陷文本分类任务。

1 继电保护系统缺陷文本分析及预处理

1.1 继电保护系统缺陷文本内容

为进一步提高继电保护系统运行可靠性, 运维人员需在日常检修和维护中对继电保护系统缺陷进行记录, 这些缺陷记录通常包含缺陷设备信息、缺陷表象描述、处理措施及缺陷级别, 其价值不容忽视。尤其是缺陷表象描述, 包含了重要的缺陷现象与可能缺陷原因, 为深入了解继电保护系统运行状态、发现潜在问题提供了宝贵的第一手资料。根据国家电网公司在《缺陷文本导则》中规定, 将继电保护系统缺陷文本缺陷级别分为 3 类, 包括“一般”、“紧急”和“严重”。本文以缺陷文本为研究对象, 通过对缺陷文本进行深度挖掘, 建立缺陷内容与缺陷级别之间的映射关系, 实现缺陷文本精准分类。部分继电保护系统缺陷文本及其缺陷级别如表 1 所示。

表 1 部分缺陷文本记录样例

Table 1 Sample of some defects text record

缺陷文本记录	缺陷级别
220kV 朝阳变电站 1 号主变 XXX-0000 变压器保护装置运行灯灭, 经检查为装置电源插件损坏, 需更换电源插件。	严重
220kV 丰台变电站 1 号主变 XXX-0000 变压器保护二次回路中央信号(监控信号)发“分合闸闭锁”, “控制回路断线”, 经检查为气压密度继电器损坏, 需更换气压密度继电器。	严重

然而, 由于缺陷文本多由运维人员手动记录, 且各运维人员书写习惯不同, 导致缺陷文本的内容并不具备统一的结构与格式。因此, 要充分利用继电保护系统缺陷文本信息, 需通过深度挖掘技术对其进行处理。

1.2 继电保护系统缺陷文本预处理

文本预处理主要包括分词、去停用词和向量化 3 个步骤。之后可将处理后的文本输入传统机器学习或深度学习模型进行特征提取与分类。

1.2.1 中文分词

中文分词是将文本转化为计算机可处理的数值表示的第一步。相较于英文，中文缺乏明显的词语分隔符。因此，中文分词旨在将连续的字串分割成有意义的词语，以便后续的文本表示和分析。

本文采用文献[25]所使用的jieba分词作为分词方法。它结合了基于词典和基于统计的两种分词方法，提供了精准、高效的中文分词解决方案。然而，由于jieba分词依靠传统大型语料库进行分词，在分割有着大量专有名词的电力行业缺陷文本时表现较差。因此本文按照词性在原词典基础上增加451个电力行业专业单词进行补充以辅助分词。新增部分词语如表2所示。

表2 部分新增文本词典

Table 2 Some new text dictionaries

词性	词典词组
名词	采样通道、保护板、二次回路、控制回路、 整定值、软压板、合闸线圈
动词	断线、失电、投退、合闸、失效
副词、形容词	异常、温高、重大、出错

1.2.2 去除停用词

停用词是指文本中出现频率高但语义价值较低的词语，属于噪声的一种。通常的去除方法是引入中文停用词表。为了提高缺陷文本分类的准确性，本文对传统的中文停用词表进行了优化。考虑到缺陷文本中的数字往往承载着重要的语义信息，本文保留了停用词表中的数字部分，以确保模型能够充分利用数字特征。

2 基于 MacBERT 模型的文本表示方法

文本表示是将文本转换为词向量的一种技术。常用的文本表示方法包括基于全局矩阵分解的潜在语义分析(latent semantic analysis, LSA)模型和基于局部上下文窗口的 Word2Vec 模型。LSA 能够捕捉全局语义，但对词汇类别区分能力较弱；Word2Vec 擅长学习局部上下文，但无法充分利用全局统计信息。本文采用 MacBERT 替代 Word2vec 和 LSA 将中文单词转化为词向量。MacBERT 是一种专门为中文自然语言处理任务而设计的一种预训练语言模型。它在 BERT 的基础上进行了改进，旨在更好地适应中文的语言特点，提升模型在各种中文 NLP 任务上的表现，相比各种传统文本表示技术有很大的

进步。其主要优点有：

1) 将掩码语言模型(masked language model, MLM)任务视为一种语言校正策略，采用相似词作为掩码字符，减小了预训练和微调阶段之间的差异，使得模型学到的表示更加鲁棒。

2) 相比 BERT 随机掩码的方式，采用全词掩码和 N-gram 掩码结合策略，更全面地覆盖文本信息，从而提高模型对语言的理解能力。

3) 相比 BERT 中使用的下句预测任务，采用 ALBERT 的句子顺序预测任务可以帮助模型更好地判断句子之间的正确顺序。

MacBERT 模型如图 1 所示。将继电保护系统缺陷文本输入到 jieba 分词工具并去停用词后得到单词 W_1 、 W_2 、 \dots 、 W_n ，以其中 W_1 为例，提取单词的语义特征、位置特征、片段特征向量分别为 W_{11} 、 W_{12} 、 W_{13} ，将得到的上述特征向量输入双向 Transformer 编码器，得到综合词向量 r_1 。

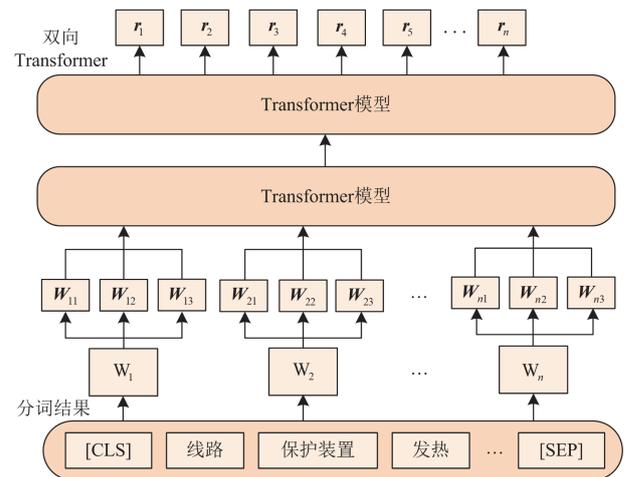


图1 MacBERT 模型

Fig. 1 Model of MacBERT

3 Linear Transformer-CNN 模型

针对 Transformer 模型训练时间过长的问题，本文首先引入线性注意力机制以提高 Transformer 模型的计算速度，之后考虑到引入线性注意力机制的 Linear Transformer 模块会丢失原有的非线性因素，导致其在最终准确率结果上表现较差。因此将善于提取局部特征的 CNN 与 Linear Transformer 通过拼接层融合，在提高分类速度的同时保证模型性能。模型具体结构如图 2 所示，主要包括 Linear Transformer 模块、多层 CNN 模块、串联拼接层、Dropout 层、SoftMax 层。

模型具体流程如下：首先将通过 MacBERT 生成的综合词向量输入到 CNN 和 Linear Transformer

模型, 分别提取局部关键特征和长距离语义特征, 并输入串联拼接层进行拼接; 最后将结果输入 SoftMax 层, 得到分类标签概率。

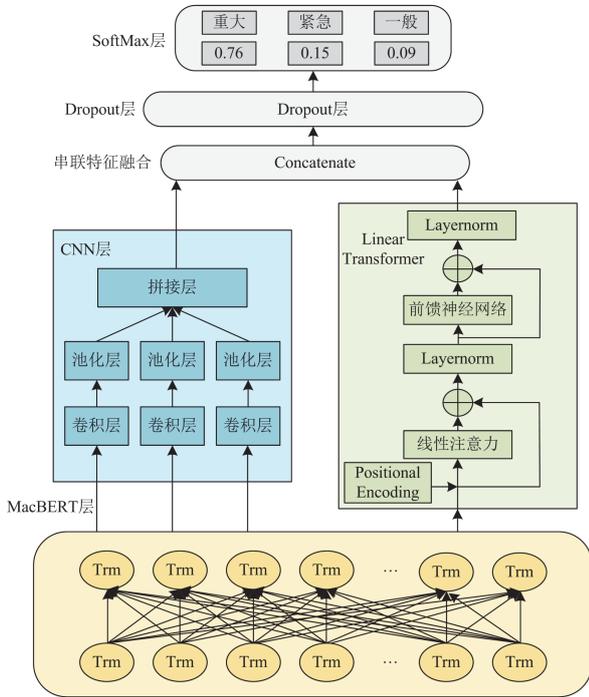


图 2 Linear Transformer-CNN 模型
Fig. 2 Linear Transformer-CNN model

3.1 Linear Transformer 模块

注意力机制是 Transformer 的核心, 其本质上是一种加权求和操作。对于序列 X 中的每个位置, 该函数会计算其与其他位置的相似度, 并根据相似度分数对其他位置的特征进行加权求和, 最终得到一个与上下文关联的特征表示。实际操作中, 输入序列 X 通过 3 个矩阵 W_Q 、 W_K 和 W_V 分别投影到查询域(Query)、键域(Key)和值域(Value), 得到相应表示矩阵 Q 、 K 和 V 。通过计算 Q 和 K 的点积, 并进

行 SoftMax 归一化, 得到注意力权重矩阵。最后, 将注意力权重矩阵与值矩阵 V 进行矩阵乘法, 得到考虑注意力机制的序列 $Attention(X)$, 计算过程如式(1)所示。

$$Attention(X) = A'_i(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

式中: $Q = XW_Q$; $K = XW_K$; $V = XW_V$; d_k 表示特征维度。

因此, 对于 $Attention(X)$ 中的某一位置 X'_i , 其注意力计算公式如式(2)所示。

$$X'_i = \frac{\sum_{j=1}^N \text{sim}(Q_i, K_j)V_j}{\sum_{j=1}^N \text{sim}(Q_i, K_j)} \quad (2)$$

式(1)和式(2)所描述的注意力机制虽然能够捕捉序列之间的长距离依赖关系, 但其在计算每一个 X'_i 时都需重新计算 sim 相似度函数, 这种二次计算复杂度限制了其在长序列上的应用, 称为 SoftMax 注意力。线性注意力机制通过引入对结合律更加友好的激活函数 $\phi(x)$, 将计算复杂度降低为线性, 从而提高了计算效率, 使得模型能够更快地处理长序列数据。具体做法如图 3 所示。

其中, 通过引入激活函数 $\phi(x) = e^{hu(x)+1}$, 可将 $\text{sim}(Q_i, K_j)$ 进行化简, 如式(3)所示。

$$\text{sim}(Q_i, K_j) = \phi(Q_i)^T \cdot \phi(K_j) \quad (3)$$

化简后的注意力机制如式(4)所示。

$$X'_i = \frac{\sum_{j=1}^N \phi(Q_i)^T \phi(K_j)V_j}{\sum_{j=1}^N \phi(Q_i)^T \phi(K_j)} \quad (4)$$

利用矩阵乘法的结合性质可进一步化简为

$$X'_i = \frac{\phi(Q_i)^T \sum_{j=1}^N \phi(K_j)V_j^T}{\phi(Q_i)^T \sum_{j=1}^N \phi(K_j)} \quad (5)$$

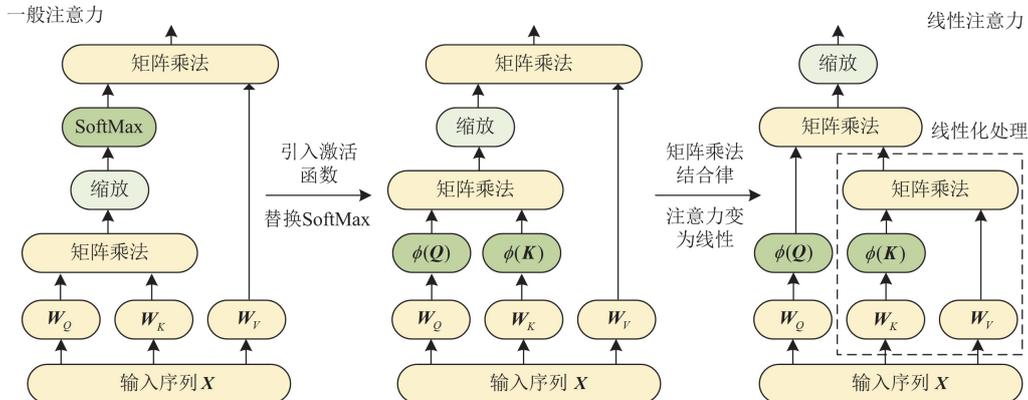


图 3 线性注意力机制

Fig. 3 Linear attention mechanism

从式(1)和式(2)可以看出, SoftMax注意力机制的计算成本为 $o(N^2)$, 其中 N 为序列长度。而线性注意力的计算成本仅为 $o(N)$, 这是因为 $\sum_{j=1}^N \phi(K_j)V_j^T$ 和 $\sum_{j=1}^N \phi(K_j)$ 与序列位置无关, 这意味着在计算每个 X'_i 时, 只需计算一次 $\sum_{j=1}^N \phi(K_j)V_j^T$ 和 $\sum_{j=1}^N \phi(K_j)$ 并为每个 $\phi(Q_i)$ 应用它们即可, 大大降低了计算复杂度。

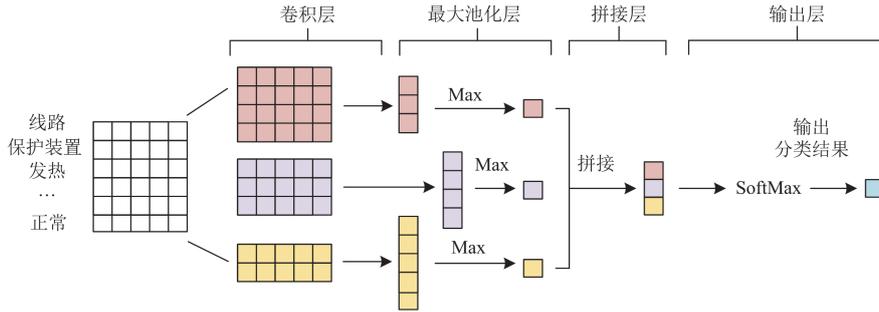


图4 TextCNN 卷积模型

Fig. 4 TextCNN convolution model

TextCNN 模块主要包括卷积层、最大池化层与拼接层, 各部分的结构及原理如下所述。

1) 卷积层

卷积层通过卷积核在词向量序列上滑动, 提取局部特征, 生成多个特征向量 X_i 。其表达式为

$$X = \{X_1, X_2, \dots, X_i\} \quad (6)$$

$$X_i = f(W \cdot Y_{i,i+h-1} + b) \quad (7)$$

式中: f 表示线性激活函数; Y 表示输入的词向量矩阵; h 为卷积核的尺寸; W 是卷积核的权重矩阵; b 是每个特征映射的偏置。

2) 最大池化层

最大池化层可将卷积层输出向量 X_i 中最重要的特征提取出来, 记为

$$X'_i = \max(X_i) \quad (8)$$

3) 拼接层

将最大池化层的不同输出向量进行拼接处理, 得到最终特征向量 X' , 记为

$$X' = \{X'_1, X'_2, \dots, X'_i\} \quad (9)$$

文本卷积是一种模拟人类阅读时关注局部词语或短语的机制。它通过在文本序列上滑动一个固定大小的窗口, 将窗口内的词向量与卷积核进行点积运算, 从而提取出局部特征。然而, 缺陷文本的表达方式往往不固定, 同一个意思可能有多种表达方式, 例如“已恢复正常”和“无异常”。为了能够有效地提取这些不同长度的特征片段, 文中采用多个

3.2 多层 TextCNN 模块

通过 3.1 节中的线性注意力方法, Transformer 模块的运行效率有了大幅的改善。然而, 由于缺乏了非线性因素, 其准确率受到影响, 稳定性相比 Transformer 较差。因此本文在 Linear Transformer 的基础上, 添加 TextCNN 模块, 对文本的局部特征进行提取, 改善模型的整体准确率。TextCNN 结构如图 4 所示。

不同大小的卷积核, 这样就可以覆盖各种可能的局部特征。

3.3 Dropout 模块

当神经网络模型的参数规模较大时, 容易发生拟合, 具体表现为模型在训练集上准确率较高, 但在测试集上准确率较低。为了缓解过拟合问题, 本文引入了 Dropout 层。Dropout 层通过随机丢弃一部分神经元, 减少模型对特定信息的依赖程度, 从而提高模型的泛化能力。这种机制使每次训练都是对一个不同的神经网络进行训练, 从而降低了模型对训练数据的过拟合程度, 提高分类准确率, 改善模型性能。

4 基于 Linear Transformer-CNN 的缺陷文本深度挖掘技术

本文提出了基于 Linear Transformer-CNN 的继电保护系统缺陷文本深度挖掘模型, 如图 5 所示。其主要步骤如下所述。

- 1) 在传统大型语料库的基础上增加电力行业专有名词, 得到适用于电力系统的专业电力词典。
- 2) 将获得的海量缺陷文本通过 jieba 分词、去停用词等操作进行数据预处理。
- 3) 采用 MacBERT 模型生成涵盖缺陷文本信息特征的综合词向量。
- 4) 搭建 Linear Transformer-CNN 模型, 并以缺

陷文本词向量作为输入、缺陷级别作为输出不断训练 Linear Transformer-CNN 模型, 优化模型各参数。

5) 进行缺陷文本深度挖掘能力测试, 在测试集中测试 Linear Transformer-CNN 模型对缺陷文本的深度挖掘能力。

6) 采用多个指标对模型的分分类效果进行综合评价, 验证模型的适用性。

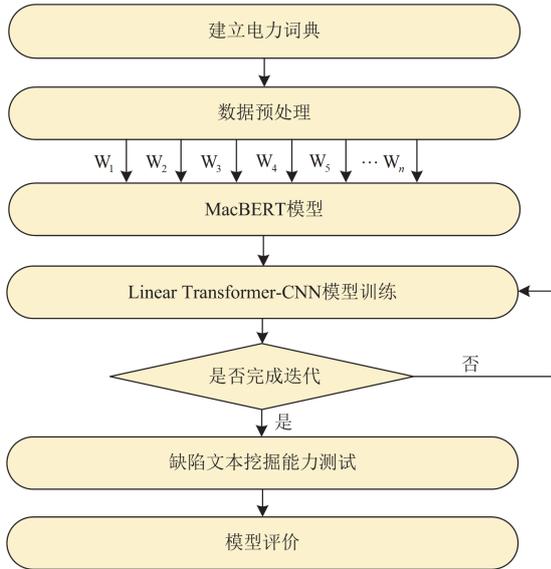


图 5 Linear Transformer-CNN 流程图

Fig. 5 Linear Transformer-CNN flowchart

5 实例分析

5.1 软硬件平台配置

深度学习模型训练对软硬件配置要求较高。本研究采用的操作系统为 Windows11, 编程平台为 Jupyter, 编程环境为 Python3.9.20, 深度学习框架为 Tensorflow2.18.0, 配置了高性能处理器以满足模型训练对计算资源的需求。具体硬件配置如表 3 所示。

表 3 具体硬件配置

Table 3 Specific hardware configurations

器件	配置
CPU	Intel(R) Core(TM) i5-13500HX
GPU	NVIDIA GeForce RTX4050
内存	16 GB

5.2 实验数据与评价指标

为验证 Linear Transformer-CNN 模型在继电保护系统的缺陷文本分类效果, 选取某电力公司近 5 年共 4815 条缺陷文本作为研究对象。这些缺陷文本包括缺陷设备信息、缺陷表象描述、处理措施、缺陷原因、缺陷级别等。缺陷级别一般有 3 种, 包括“一般”、“紧急”和“严重”。本文采用 random_split

函数将原始数据按 8:2 随机划分为训练集和测试集。各类数据数目如表 4 所示。

表 4 缺陷文本数据量

Table 4 Amount of defective text data

数据集	一般	严重	紧急	合计
训练集	2012	912	928	3852
测试集	492	236	235	963
合计	2504	1148	1163	4815

为了全面评估模型的性能, 采用精确率 P 、准确率 A 、训练耗时 T_{train} 、测试耗时 T_{test} 4 项指标对模型进行多维度评价, 并与其他模型进行对比。具体计算公式为

$$A = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (10)$$

$$P = \frac{T_p}{T_p + F_p} \quad (11)$$

式中: T_p 、 T_N 、 F_p 、 F_N 为混淆矩阵元素。混淆矩阵如表 5 所示。

表 5 混淆矩阵

Table 5 Confusion matrix

预测类别	实际类别	
	实际为 A 类事件	实际为 B 类事件
预测为 A 类事件	T_p	F_p
预测为 B 类事件	F_N	T_N

继电保护系统缺陷文本分类问题为三分类问题。其中: 准确率 A 指所有缺陷文本缺陷等级分类中正确的比例; 精确率 P 指在预测为一般缺陷中实际为一般缺陷的占比。

5.3 词向量可视化处理

继电保护系统缺陷文本经 jieba 分词、去停用词、MacBERT 等操作得到词向量后, 每个词被表示为一个 64 维的高维向量。为了更直观地展示这些高维向量特征关系, 采用均匀流形近似与投影(uniform manifold approximation and projection, UMAP)算法将其降至三维空间。

UMAP 算法是一种基于图论和流形学习的降维算法, 可以将高维数据映射到低维空间, 以便于看到词语之间的特征联系。对 MacBERT 输出的词向量进行 UMAP 算法降维分析, 结果如表 6 和图 6 所示。

以“CPU 插件”、“cpu 插件”和“CPU 模块”举例, 从图中可以看出三者距离很近, 表明它们在语义上的关联性很大, 这与实际情况是一致的。证明 MacBERT 生成的词向量包含准确的词语特征。

此外, 为进一步验证 MacBERT 模型的有效性, 分别使用 MacBERT + Linear Transformer-CNN 模型

与 BERT + Linear Transformer-CNN 模型进行缺陷文本分类实验, 实验结果如图 7 所示。

表 6 部分缺陷文本分词结果

Table 6 Partial defective text tokenization results

序号	缺陷文本分词前	缺陷文本分词后
1	220kV 朝阳变电站 1 号主变 XXX-0000 变压器保护装置运行灯灭, 经检查为装置电源插件损坏, 需更换电源插件。	'220kV','变电站','1 号主变','XXX0000','变压器','保护装置','运行','灯灭','检查','装置','电源','插件','损坏','更换','电源','插件'
2	220kV 丰台变电站 1 号主变 XXX-0000 变压器保护二次回路中央信号(监控信号)发“分合闸闭锁”, “控制回路断线”, 经检查为气压密度继电器损坏, 需更换气压密度继电器。	'220kV','变电站','1 号主变','XXX0000','变压器','保护','二次','回路','中央','信号','监控','信号','发','分','合闸','闭锁','控制','回路','断线','检查','气压','密度','继电器','损坏','更换','气压','密度','继电器'

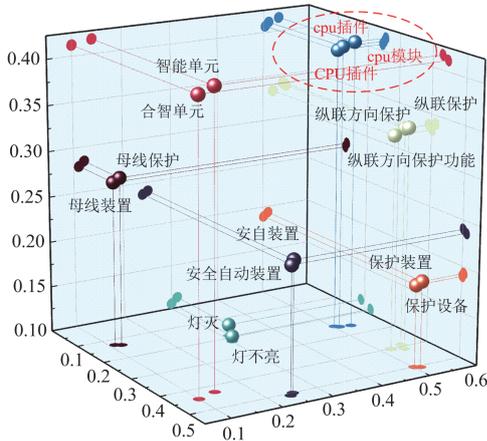


图 6 部分词向量可视化显示

Fig. 6 Word vector visualization

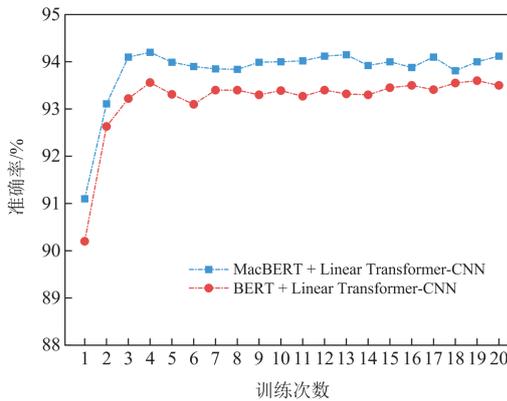


图 7 MacBERT 模型有效性验证

Fig. 7 Validation of the validity of the MacBERT model

从图 7 中可以看出, 采用 MacBERT 模型生成词向量进而实现缺陷文本深度挖掘的效果更好, 其

准确率明显优于 BERT 模型。

5.4 模型性能分析

为验证 Linear Transformer-CNN 模型的改进分类效果, 分别使用 Linear Transformer-CNN、Linear Transformer、Transformer、Transformer-CNN 和 TextCNN 模型进行缺陷文本分类实验。为保证实验结果的有效性, 所有实验模型均采用 MacBERT 对单词进行词向量表示。Linear Transformer-CNN 模型参数如表 7 所示, 实验结果如图 8 和表 8 所示。

表 7 Linear Transformer-CNN 模型参数

Table 7 Linear Transformer-CNN model parameters

训练参数	设置值	参数含义
Epoch	20	训练次数
Learning_rate	0.001	学习率
Dropout_rate	0.1	丢弃率
Max_length	150	MacBERT 最大读取序列长度
Size	3,4,5	卷积核尺寸
Number	64	卷积核数量

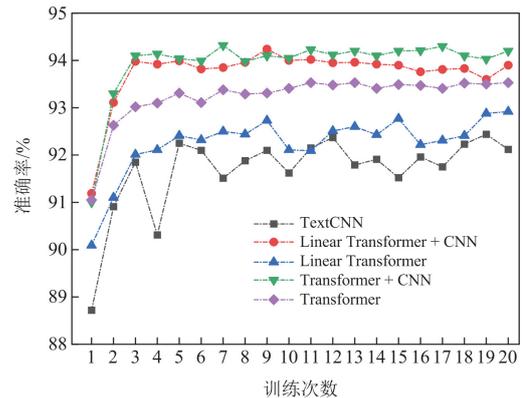


图 8 不同模型的准确度对比结果

Fig. 8 Accuracy comparison results of different models

表 8 不同模型的实验结果

Table 8 Experimental results of different models

模型	准确率/%	精确率/%	训练时间/s	测试耗时/s
TextCNN	92.44	92.69	25.97	0.19
Linear Transformer	92.92	93.31	33.44	0.31
Transformer	93.53	93.97	54.85	0.56
Transformer-CNN	94.32	94.89	82.15	0.76
Linear	94.24	94.51	62.34	0.51
Transformer-CNN				

由表 8 可知, Linear Transformer-CNN 模型的分类准确率相比于 TextCNN、Linear Transformer、分别提高 1.8%和 1.32%。Linear Transformer 相比于 Transformer, 其计算速率提升 1.64 倍, 然而, 由于缺少非线性因素, 其准确率有所下降, 且稳定性较

差, 在文本分类准确率上表现糟糕; TextCNN 模型由于无法把握长距离语义信息, 在文本分类准确度上表现同样不够优秀。在训练时间上, Linear Transformer-CNN 模型相比 Transformer-CNN 模型缩短 19.81 s, 这是由于 Transformer-CNN 模型所需参数过多, 且计算具有二次复杂度, 在对海量缺陷文本进行处理时, 其缺陷暴露更加明显。综合来看, 本文提出的 Linear Transformer-CNN 缺陷文本分类模型综合性能更优, 能够完成继电保护系统缺陷文本深度挖掘工作。

5.5 与机器学习模型性能对比

为对比 Linear Transformer-CNN 与其他机器学习模型的性能, 采用朴素贝叶斯、Gradient Boosting、K 近邻和决策树 4 种算法分别进行继电保护系统缺陷文本分类, 结果如图 9 所示。

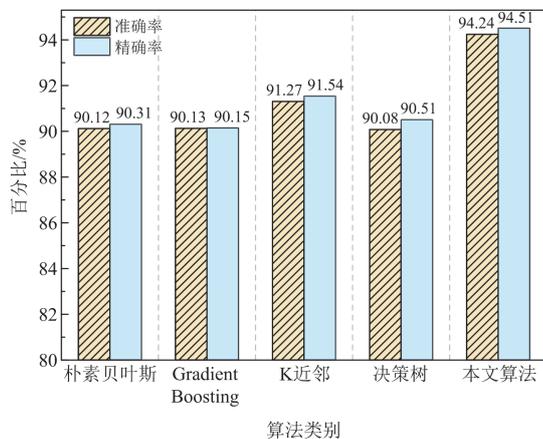


图 9 与机器学习对比结果

Fig. 9 Comparison results with machine learning

由图 9 可知, 本文所提方法相比于传统的机器学习算法在准确率和精确率上都有大幅提高, 相比于其中表现最优的 K 近邻算法, 其也在精确率和准确率上均高出了 2.97%。

5.6 与深度学习模型性能对比

为对比 Linear Transformer-CNN 与其他深度学习模型的性能, 采用 TextCNN、LSTM、BiLSTM 和 Transformer 4 种算法分别进行继电保护系统缺陷文本分类, 结果如图 10 所示。

由图 10 可知, 本文方法性能相比于其他深度学习算法都有一定程度的提升, 除了上文已经提过的 TextCNN 和 Transformer, 本文算法相比于 LSTM 和 BiLSTM 算法, 准确率分别提升了 1.35% 和 1.12%。综合整体对比结果, 本文采用的 Linear Transformer-CNN 模型性能明显优于机器学习算法和深度学习算法。其优势在于将 Linear Transformer

提供的快速、准确长语义特征识别能力与 TextCNN 模型提供的局部特征提取能力进行融合, 实现了对继电保护系统缺陷文本的快速精准分类。

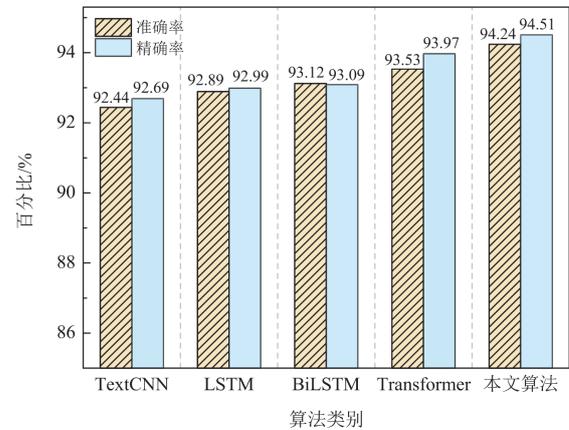


图 10 与深度学习对比结果

Fig. 10 Comparison results with deep learning

6 结论

本文提出了一种基于 Linear Transformer-CNN 模型的继电保护系统缺陷文本深度挖掘方法, 旨在结合 Linear Transformer 对缺陷文本长距离语义特征能够快速提取的优点和 CNN 模型较好的局部特征提取优点, 为继电保护系统缺陷文本的深度挖掘问题提供一种新的思路。

1) 采用计算机领域的文本挖掘流程, 利用 MacBERT 模型实现继电保护系统缺陷文本向量化。通过可视化处理将词向量映射到三维空间, 结果表明, 所生成的词向量特征关系清晰, 证明该模型具有良好的文本表示能力。

2) 在 Transformer 模型的基础上引入线性注意力机制得到 Linear Transformer 模型, 有效地减轻了模型的计算复杂度。实验证明, Linear Transformer 相比于 Transformer 模型, 所需训练时间缩短 1.64 倍, 然而其准确率过低, 仅有 92.92%, 不适合缺陷文本深度挖掘工作。

3) 在 Linear Transformer 模型的基础上, 为了弥补其丢失非线性特征引起的最终准确率下降的问题, 引入 CNN 模块以提升整体分类准确率。经大量实验结果表明, 本文所提 Linear Transformer-CNN 方法达到 94.24% 的高准确率, 具有良好的应用价值。

参考文献

- [1] 许尧, 马欢, 许昆鹏, 等. 智能变电站继电保护智能运维系统自动配置技术研究[J]. 电力系统保护与控制, 2022, 50(11): 160-168.

- XU Yao, MA Huan, XU Chanpeng, et al. Self-configuration technology of an intelligent operation and maintenance system of intelligent substation relay protection[J]. *Power System Protection and Control*, 2022, 50(11): 160-168.
- [2] 章耀耀, 侯勇, 叶海, 等. 继电保护数字孪生技术应用展望[J]. *电力系统保护与控制*, 2023, 51(12): 178-187.
ZHANG Yaoyao, HOU Yong, YE Hai, et al. Application prospect for relay protection digital twin technology[J]. *Power System Protection and Control*, 2023, 51(12): 178-187.
- [3] 王增平, 林一峰, 王彤, 等. 电力系统继电保护与安全控制面临的挑战与应对措施[J]. *电力系统保护与控制*, 2023, 51(6): 10-20.
WANG Zengping, LIN Yifeng, WANG Tong, et al. Challenges and countermeasures to power system relay protection and safety control[J]. *Power System Protection and Control*, 2023, 51(6): 10-20.
- [4] 徐丙垠. 现代配电网继电保护技术及其发展[J]. *供用电*, 2024, 41(8): 65-74.
XU Bingyin. Modern distribution systems relay protection technology and its developments[J]. *Distribution & Utilization*, 2024, 41(8): 65-74.
- [5] YAO Zhiqing, LI Danyang, LI Zhiyong, et al. Relay protection mirror operation technology based on digital twin[J]. *Protection and Control of Modern Power Systems*, 2023, 8(4): 1-14.
- [6] 于晓军, 刘志远, 吴建云, 等. 基于无线同步信息的继电保护分布式智能测试系统[J]. *电力系统保护与控制*, 2022, 50(11): 121-129.
YU Xiaojun, LIU Zhiyuan, WU Jianyun, et al. A distributed intelligent testing system of relay protection based on wireless synchronization information[J]. *Power System Protection and Control*, 2022, 50(11): 121-129.
- [7] 刘中硕, 郑少明, 陶畅, 等. 继电保护装置缺陷文本专业词典构建及其语言特性分析[J]. *中国电力*, 2023, 56(7): 146-155.
LIU Zhongshuo, ZHENG Shaoming, TAO Chang, et al. The construction of the professional dictionary of relay protection defect text in a regional power grid and its natural language characteristics analysis[J]. *Electric Power*, 2023, 56(7): 146-155.
- [8] 陶畅. 继电保护装置缺陷文本数据挖掘技术及其应用[D]. 北京: 华北电力大学, 2023.
TAO Chang. Relay protection devices defect text data mining technology and application[D]. Beijing: North China Electric Power University, 2023.
- [9] 戴志辉, 张富泽, 张近月, 等. 基于 MacBERT-BiLSTM-CRF 模型的继电保护装置缺陷知识图谱构建方法[J]. *电力系统保护与控制*, 2024, 52(20): 131-143.
DAI Zhihui, ZHANG Fuze, ZHANG Jinyue, et al. Construction method of a defect knowledge map of a relay protection device based on a MacBERT-BiLSTM-CRF model[J]. *Power System Protection and Control*, 2024, 52(20): 131-143.
- [10] 张磐, 郑悦, 李海龙, 等. 配电网电力设备缺陷文本智能辨识运维综述[J]. *电力建设*, 2022, 43(5): 90-99.
ZHANG Pan, ZHENG Yue, LI Hailong, et al. Overview on intelligent text identification and maintenance of power equipment defects in distribution network[J]. *Electric Power Construction*, 2022, 43(5): 90-99.
- [11] 李元, 李睿, 林金山, 等. 基于字词混用集成模型的电力变压器缺陷记录文本挖掘方法[J]. *电力工程技术*, 2024, 43(6): 153-162.
LI Yuan, LI Rui, LIN Jinshan, et al. Character-word level ensemble integrated model for power transformer defect recording text mining method[J]. *Electric Power Engineering Technology*, 2024, 43(6): 153-162.
- [12] 邵冠宇, 王慧芳, 何奔腾. 电网设备缺陷文本的质量评价与提升方法[J]. *电网技术*, 2019, 43(4): 1472-1479.
SHAO Guanyu, WANG Huifang, HE Benteng. Quality assessment and improvement method for power grid equipment defect text[J]. *Power System Technology*, 2019, 43(4): 1472-1479.
- [13] 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.
- [14] 廖才波, 黄智勇, 杨金鑫, 等. 基于缺陷文本识别的变压器风险评估及辅助检修决策方法[J]. *高电压技术*, 2024, 50(7): 2931-2941.
LIAO Caibo, HUANG Zhiyong, YANG Jinxin, et al. Risk assessment and auxiliary maintenance decision method of transformer based on defect text recognition[J]. *High Voltage Engineering*, 2024, 50(7): 2931-2941.
- [15] KIM Y. Convolutional neural networks for sentence classification[C] // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, October 25-29, 2014, Doha, Qatar: 1746-1751.
- [16] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C] // *31st Annual Conference on Neural Information Processing Systems*, December 4-6, 2017,

- California, USA: 6000-6010.
- [18] 马润泽, 王龙响, 余佳文, 等. 考虑历史缺陷文本信息的断路器状态评价研究[J]. 机电工程, 2015, 32(10): 1375-1379.
MA Runze, WANG Longxiang, YU Jiawen, et al. Circuit breakers condition evaluation considering the information in historical defect texts[J]. Journal of Mechanical & Electrical Engineering, 2015, 32(10): 1375-1379.
- [19] 刘梓权, 王慧芳, 曹靖, 等. 基于卷积神经网络的电力设备缺陷文本分类模型研究[J]. 电网技术, 2018, 42(2): 644-650.
LIU Ziquan, WANG Huifang, CAO Jing, et al. A classification model of power equipment defect texts based on convolutional neural network[J]. Power System Technology, 2018, 42(2): 644-650.
- [20] 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.
- [21] 冯斌, 张又文, 唐昕, 等. 基于 BiLSTM-Attention 神经网络的电力设备缺陷文本挖掘[J]. 中国电机工程学报, 2020, 40(增刊 1): 1-10.
FENG Bin, ZHANG Youwen, TANG Xin, et al. Power equipment defect record text mining based on BiLSTM-Attention neural network[J]. Proceedings of the CSEE, 2020, 40(S1): 1-10.
- [22] 张中文, 吐松江·卡日, 张紫薇, 等. 基于双分支特征融合的电力设备缺陷文本挖掘方法[J]. 高压电器, 2024, 60(6): 188-196.
ZHANG Zhongwen, TUSONGJIANG Kari, ZHANG Ziwei, et al. Text mining method for power equipment defects based on two-branch feature fusion[J]. High Voltage Apparatus, 2024, 60(6): 188-196.
- [23] DEVLIN J, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, June 2-7, 2019, Stroudsburg, USA: 4171-4186.
- [24] 夏飞, 陈帅琦, 华珉, 等. 基于改进 BERT 的电力领域中文分词方法[J]. 计算机应用, 2023, 1(1): 1-10.
XIA Fei, CHEN Shuaiqi, HUA Min, et al. Chinese word segmentation method in electric power domain based on improved BERT[J]. Journal of Computer Applications, 2023, 1(1): 1-10.
- [25] 虞佳淼, 王慧芳, 张亦翔, 等. 基于 BERT 的电网现场作业风险自动评级方法[J]. 电网技术, 2023, 47(11): 4746-4754.
YU Jiamiao, WANG Huifang, ZHANG Yixiang, et al. Automatic risk rating method for power grid field operation based on BERT[J]. Power System Technology, 2023, 47(11): 4746-4754.

收稿日期: 2024-11-12; 修回日期: 2025-02-17

作者简介:

戴志辉(1980—), 男, 博士, 教授, 研究方向为电力系统保护与控制; E-mail: daihuadian@163.com

贺欲飞(2001—), 男, 硕士研究生, 研究方向为电力系统保护与安全防御。E-mail: hencepu@163.com

(编辑 魏小丽)