

DOI: 10.19783/j.cnki.pspc.250335

# 基于 Sentence-MacBERT 模型的虚端子自动连接方法

戴志辉, 贺欲飞, 张洪嘉

(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:** 鉴于智能变电站待连接的虚端子数量庞大, 传统的连接方法常存在识别效率低、校核工作量大、连接结果不准确等问题。提出一种基于掩码纠错型双向编码器句子嵌入模型(sentence-masked language model as correction bidirectional encoder representations from transformer, Sentence-MacBERT)的虚端子自动连接方法。首先, 提取实现虚端子自动连接所需关键信息并进行预处理。其次, 构建 Sentence-MacBERT 虚端子自动连接模型并进行训练, 得到最优模型。最后, 将预处理后的短地址和中文描述分别输入到该模型中, 得到综合句向量并进行余弦相似度匹配, 完成智能变电站虚端子自动连接。结果表明, 相比于传统的虚端子自动连接方法, 该方法的连接效率更高, 且准确率达到 94.38%, 实现了虚端子的准确连接。

**关键词:** 虚端子; 智能变电站; Sentence-MacBERT; 深度学习; 文本匹配

## An automatic virtual terminal connection method based on the Sentence-MacBERT model

DAI Zhihui, HE Yufei, ZHANG Hongjia

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

**Abstract:** In view of the large number of virtual terminals to be connected in smart substations, traditional connection methods often suffer from low identification efficiency, large verification workload, and inaccurate connection results. To address these issues, an automatic virtual terminal connection method based on the sentence-masked language model as correction bidirectional encoder representations from transformer (Sentence-MacBERT) model is proposed. First, the key information required for the automatic virtual terminal connection is extracted and preprocessed. Second, the Sentence-MacBERT virtual terminal automatic connection model is constructed and trained to obtain the optimal model. Finally, the preprocessed short addresses and Chinese descriptions are separately input into the model to obtain the comprehensive sentence embeddings, and cosine similarity matching is carried out to complete the automatic connection of virtual terminals in smart substations. The results show that, compared with traditional virtual terminal automatic connection methods, the proposed method achieves higher connection efficiency, with an accuracy of 94.38%, enabling accurate virtual terminal connection.

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

**Key words:** virtual terminal; smart substation; Sentence-MacBERT; deep learning; text matching

## 0 引言

在传统变电站中, 端子间的连接主要通过二次电缆实现, 这些电缆携带了大量的电气信息。然而, 由于各电气信息需通过独立的电缆进行传输, 导致变电站中电缆数量庞大, 维护成本极高<sup>[1-4]</sup>。因此, 设计人员在智能变电站中采用光纤通过通信网络传输数字信号以取代传统变电站中的电气信号<sup>[5-9]</sup>。

由于单根光纤往往承载多路数据, 需借助虚端子确保数据传输的正确性<sup>[10-12]</sup>。

目前, 国内设计院在智能变电站中主要通过人工对虚端子进行连接<sup>[13]</sup>。设计人员往往通过厂家提供的智能电子装置(intelligent electronic device, IED)能力描述文件手动创建虚端子表以反映装置间的虚端子连接关系。这种方式在效率和准确率上表现较差。此外, 当变电站需要进行改扩建时, 模型文件的变更会导致虚端子连接关系全部丢失, 须再次通过人工进行连接, 大幅增加了设计人员的工作量。因此,

基金项目: 国家自然科学基金项目资助(51877084)

亟需研究虚端子自动连接技术, 以提高连接效率<sup>[14]</sup>。

国内目前针对虚端子自动连接的研究较少。文献[15]提出了一种基于关键字匹配的方法, 实现了虚端子的连接, 然而其关键字仍依赖人工经验进行选取, 存在大量主观因素, 对连接准确度和效率存在影响。文献[16]通过对二次设备自动关联理论和需求进行分析, 提出统一的虚端子格式标准, 并利用关键字编码, 实现虚端子表的自动生成。文献[17]构建了智能变电站虚端子辅助自动设计软件系统, 提高了设计的效率与准确性。文献[18]通过编辑距离定义了虚端子间的相似度, 并通过粒子群优化算法得到距离权重向量, 最终实现自动连接; 然而这种方法计算过程较为复杂, 限制了其连接效率。文献[19]提出了一种基于改进 Dice 距离的虚端子连接准确性判断方法, 该方法能有效找出连接错误的虚端子并进行修正, 然而在面对海量虚端子时仍显不足。文献[20]构建了智能变电站虚回路监视技术体系, 为虚回路的可视化展示提供数据基础。文献[13]提出了一种基于烟花算法计算距离权重相似度的虚端子自动连接方法, 实现了待匹配 IED 设备虚端子的自动相连。

除此之外, 谷歌在 2019 年提出的基于 Transformer 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突破<sup>[21]</sup>。相较于 Word2Vec, BERT 及其衍生模型凭借双向编码结构, 在文本处理方面展现出卓越性能<sup>[22-23]</sup>。由于虚端子表中同样存在大量文本, 这一模型也为虚端子的自动连接带来了新的可能。现有的处理方法已无法满足海量虚端子文本分析的需要, 通过引用并改进 BERT 模型, 可大幅提升虚端子文本信息处理效率。

针对上述问题, 本文提出了一种基于掩码纠错型双向编码器句子嵌入模型(sentence-masked language model as correction bidirectional encoder representations from transformer, Sentence-MacBERT)的虚端子自动

连接方法。首先, 提取虚端子连接所需关键信息并进行预处理。然后, 将预处理后的关键信息输入到 Sentence-MacBERT 模型中生成对应综合句向量。最后, 通过计算各向量的余弦相似度得到连接结果。结果表明, 基于 Sentence-MacBERT 模型的虚端子自动连接方法准确率高、效果好, 能胜任自动连接任务。

## 1 虚端子的格式与特点

在智能变电站中, 虚端子不是物理存在的实体, 它代表了不同设备之间逻辑上的连接关系。按照所传输信号协议的不同, 可分为 SV 虚端子和 GOOSE 虚端子。在国标 GB/T 36273-2018 中规定, SV 虚端子所传信息主要包括电流通路、电压通道及固有延时等, GOOSE 虚端子所传信息主要包括开关量输入、跳闸出口、信号开出、告警、联闭锁等。除此之外, 根据信号传输方向的不同, 又分为输出虚端子和输入虚端子。因此, 虚端子在智能变电站中共分 4 类, 包括 GOOSE 输入虚端子、GOOSE 输出虚端子、SV 输出虚端子、SV 输入虚端子。

IEC61850 标准为智能变电站二次系统建立了一套统一的模型和通信协议。该标准通过对数据的名称、功能、结构进行严格规定, 实现了不同厂商设备之间的数据共享和互操作。该标准具有较强的层次性, 从上到下分为物理设备 PD、逻辑设备 LD、逻辑节点 LN、数据对象 DO 和数据属性 DA 等层次, 并为每个数据元素分配唯一标识。这种层次化结构使得模型具有高度的灵活性和扩展性, 能够适应智能电网不断发展的需求。IEC61850 标准要求 GOOSE 虚端子的标准数据格式应配置到 DA 层次, SV 虚端子配置到 DO 层次, 其标准格式分别为 LD/LN.DO.DA 和 LD/LN.DO。除此之外, 虚端子还需增加一段中文短文本描述以表明其物理含义, 某 110 kV 变电站中部分虚端子如表 1 所示。

表 1 部分虚端子表

Table 1 Some imaginary terminals

IED 设备名称	虚端子中文描述	虚端子短地址
110 kV 分段二 A 套合智一体装置	保护电路 A 相 1	IMF1102A.MUSV01/TCTR1.AMP
110 kV 分段二 A 套合智一体装置	保护电路 A 相 2	IMF1102A.MUSV01/TCTR1.AMPChB
110 kV 分段二 A 套合智一体装置	保护电路 B 相 1	IMF1102A.MUSV01/TCTR2.AMP
110 kV 分段二 A 套合智一体装置	保护电路 B 相 2	IMF1102A.MUSV01/TCTR2.AMPChB
110 kV 分段二 A 套合智一体装置	保护电路 C 相 1	MF1102A.MUSV01/TCTR3.AMP
110 kV 分段二 A 套合智一体装置	保护电路 C 相 2	IMF1102A.MUSV01/TCTR3.AMPChB
#2 主变测控装置	刀闸 3 分闸操作	CT1102.PI/QCSWI8.OpOpn.general
#2 主变测控装置	刀闸 3 合闸操作	CT1102.PI/QCSWI8.OpCls.general
#2 主变测控装置	接地刀闸 4 分闸操作	CT1102.PI/QCSWI19.OpOpn.general
#2 主变测控装置	接地刀闸 4 合闸操作	CT1102.PI/QCSWI19.OpCls.general

从表 1 中可知，虚端子连接所需关键信息主要为短地址和中文描述。因此，要想实现虚端子自动连接，须对短地址和中文描述进行分析处理。

## 2 虚端子短地址处理方法

### 2.1 虚端子短地址预处理

虚端子短地址是一串字符串，为提高虚端子连接准确度，需对虚端子短地址进行预处理，主要包括以下几步。

#### 1) 对虚端子短地址进行分词

在 IEC61850 中已对虚端子短地址格式进行了严格要求，SV 和 GOOSE 虚端子地址格式分别为 LD/LN.DO 和 LD/LN.DO.DA，只需要按照既有层次进行分词即可。

#### 2) 去除多余信息

在对某 IED 设备虚端子进行连接时，存在着大量冗余信息，这些冗余信息不仅会占用内存，还会对连接准确性造成影响。例如在对某合智一体装置虚端子进行连接时，由于其逻辑设备均为 MUSV01，此时可利用 MATLAB 中函数将其去除。部分虚端子预处理结果如表 2 所示。

表 2 虚端子短地址处理结果

Table 2 Processing results of short addresses of imaginary terminals

短地址处理前	短地址处理后
MUSV01/SVINTVTR1.Vol.instMag.i	“TVTR1” “Vol”
MUSV01/SVINTVTR1.VolChB.instMag.i	“TVTR1” “VolChB”
MUSV01/SVINTVTR2.Vol.instMag.i	“TVTR2” “Vol”
MUSV01/SVINTVTR2.VolChB.instMag.i	“TVTR2” “VolChB”
MUSV01/SVINTVTR3.Vol.instMag.i	“TVTR3” “Vol”
MUSV01/SVINTVTR3.VolChB.instMag.i	“TVTR3” “VolChB”
MUSV01/SVINTVTR4.Vol.instMag.i	“TVTR4” “Vol”
MUSV01/SVINTVTR4.VolChB.instMag.i	“TVTR4” “VolChB”
MUSV01/SVINTVTR5.Vol.instMag.i	“TVTR5” “Vol”
MUSV01/SVINTVTR5.VolChB.instMag.i	“TVTR5” “VolChB”

### 2.2 BERT 预训练模型

文本表示是将文本映射为计算机可理解的数值表示过程，其核心在于将词语进行向量化。传统方法如基于特定编码方式的 One-Hot 模型和基于局部上下文窗口的 Word2Vec 模型在文本表示上都存在一定缺点。例如经 One-Hot 模型处理后的单词会得到一个高维但稀疏的词向量，占用大量的存储和计算空间。Word2Vec 模型无法考虑全文信息且对词语的语义信息理解有限。因此，本文采用 BERT 预训练模型替代 Word2vec 和 One-Hot 模型，将上述预处理后的短地址单词转化为词向量。BERT 模型结

构如图 1 所示。

在图 1 中，[CLS]和[SEP]标记分别作为输入序列的起始和结束符号。将单词输入到 BERT 模型中经双向 Transformer 编码器特征提取，可得到包含丰富词语信息的向量表示。BERT 模型的预训练阶段主要包含两个任务：掩码语言模型(masked language model, MLM)任务和下句预测(next sentence prediction, NSP)任务。掩码语言模型任务即随机遮盖部分输入词，并训练模型预测被遮盖的词，从而学习上下文语义；下句预测任务通过学习句子间的特征关系来预测两句子在语义上是否连贯，从而提升模型对句子的理解。通过这两个任务的联合训练，BERT 模型能够学习到更深层次的语言表示。

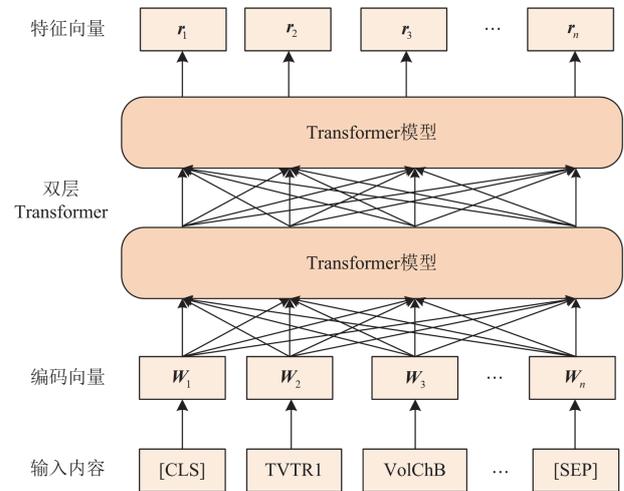


图 1 BERT 预训练模型

Fig. 1 BERT pre-trained model

## 3 虚端子中文描述处理方法

目前，虚端子中文描述常用的处理方法是通过对计算编辑距离来表示文本的相似程度，从而判断两句文本描述所对应的虚端子是否匹配。然而，这种距离计算只考虑了两句文本表面上的相似程度，没有深度考虑其实际的语义信息。事实上，对于虚端子中文描述的匹配问题，其语义是否匹配是研究重点。为此，本文利用自然语言处理方法，对虚端子中文描述进行处理分析，从而为实现虚端子自动连接提供基础。

### 3.1 基于 Pkuseg 工具的虚端子中文描述预处理

中文分词是将中文文本转化为计算机可处理的数值表示第一步。目前常用的分词工具是 jieba 和 Pkuseg 分词。相较于 Jieba 分词工具，Pkuseg 的显著特点在于支持多领域分词，且能自训练分词模型。通过为不同领域提供预训练模型，Pkuseg 能够根据文本的领域特点选择最适合的模型进行分词，从而

显著提升分词的准确性。此外, 用户可以根据自身需求, 仅提供已分词的领域数据集, 即可训练出定制化的分词模型。本文将已分词好的虚端子中文描述构成自定义词典, 测试并验证 Pkuseg 分词工具的分词能力, 结果如表 3 所示。

表 3 Pkuseg 分词工具结果

Table 3 Results of the Pkuseg word segmentation tool

处理前中文描述	Pkuseg 分词处理后	引入自定义词典的 Pkuseg 分词处理后
110 kV II 母通道延迟	110 kV II/母/通道/延迟	110 kV/II 母/通道/延迟
110 kV III 母 PT 合并单元	110 kV III/母/PT/合并/单元	110 kV/III 母/PT/合并单元
母线 2A 相保护电压	母线/2/A/相/保护/电压	母线 2/A 相/保护/电压
III 段零序电压采样值	III/段/零/序/电压/采样/值	III 段/零序/电压/采样/值
复采 II 侧 A 相电流	复采/II/侧/A/相/电流	复采/II 侧/A 相/电流

### 3.2 MacBERT 中文文本表示模型

MacBERT 模型专门针对中文设计, 能够更好地捕捉中文的语义和语法信息。相比于 BERT 模型, MacBERT 做了大量改进。首先, MacBERT 将掩码语言模型任务视为一种语言校正方式, 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; 其次, 相比 BERT 随机掩码的方式, MacBERT 采用全词掩码和 N-gram 掩码结合的策略, 更全面地覆盖文本信息, 从而提高模型对语言的理解能力。

MacBERT 模型如图 2 所示, 将虚端子中文描述输入到模型中得到单词  $W_1$ 、 $W_2$ 、 $W_3$ , 以其中  $W_1$  为例, 提取单词的标记向量、片段向量和位置向量分别为  $W_{11}$ 、 $W_{12}$ 、 $W_{13}$ , 其中, 标记特征向量代表文本每个字符的原始向量表示; 片段特征向量用于区分输入文本中的不同句子; 位置特征向量则对字符在文本中的位置信息进行编码, 生成相应的特征向量。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  $r_1$ 。

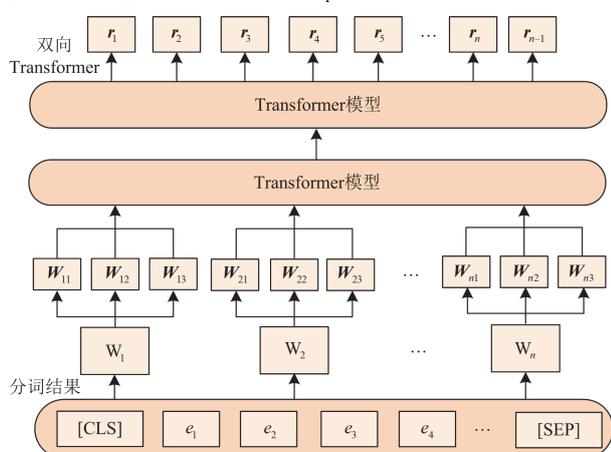


图 2 MacBERT 模型

Fig. 2 Model of MacBERT

从表 3 中可以看出, 引入自定义词典的 Pkuseg 分词工具的表现较好, 尤其是在数字和专有名词的划分上, 引入自定义词典的 Pkuseg 分词能够更好地捕捉数字与专有名词特征, 做到对中文描述的准确划分。

## 4 基于 Sentence-MacBERT 的虚端子自动连接模型

### 4.1 Sentence-BERT 模型

Sentence-BERT 模型在 BERT 模型的输出结果后加入了池化层, 其目的在于使 BERT 生成的多个词向量转换为一个固定长度的句向量, 以便后续的余弦相似度计算。常用的池化策略有 3 种, 包括最大池化、平均池化和 CLS 池化。本文采用平均池化策略, 即将所有单词的输出向量取平均值作为待求句子向量。其具体结构如图 3 所示。

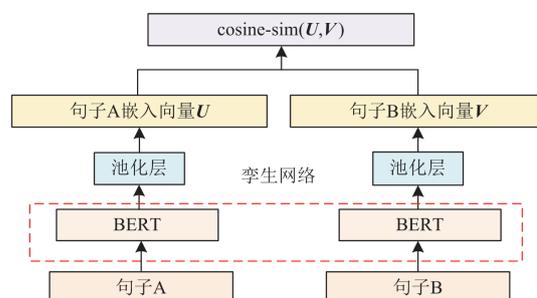


图 3 Sentence-BERT 模型

Fig. 3 Sentence-BERT model

除此之外, Sentence-BERT 模型利用孪生网络对模型进行微调, 使生成的句向量具有更准确的语义, 增加其对词语的特征提取能力, 提高运行效率。

### 4.2 基于 Sentence-MacBERT 的虚端子连接模型

虚端子的自动连接重点在于对其短地址和中文描述的处理, 考虑到 MacBERT 模型对中文文本有着更好的处理效果, 本文分别用 MacBERT 和 BERT 模型对中文文本和短地址进行处理。为了方便统一名称, 本文将该模型命名为 Sentence-MacBERT 模型。构建得到基于 Sentence-MacBERT 模型的虚端子自动连接模型, 具体结构如图 4 所示。

从图 4 中可以看出，将预处理后的虚端子短地址和文本描述分别输入到 BERT 和 MacBERT 模型后经池化层生成句向量，再通过串联拼接层将生成的句向量进行拼接得到综合句向量，最后通过计算综合句向量的余弦相似度衡量两虚端子的匹配程度，如式(1)所示。

$$\cos \theta = \frac{U \cdot V}{|U||V|} \quad (1)$$

式中： $U$ 、 $V$  分别代表两句子的综合句向量。

余弦相似度大小反映了虚端子连接情况，越高则代表输入虚端子匹配度越高，本文最终选取余弦相似度最高的虚端子匹配项作为自动连接结果。

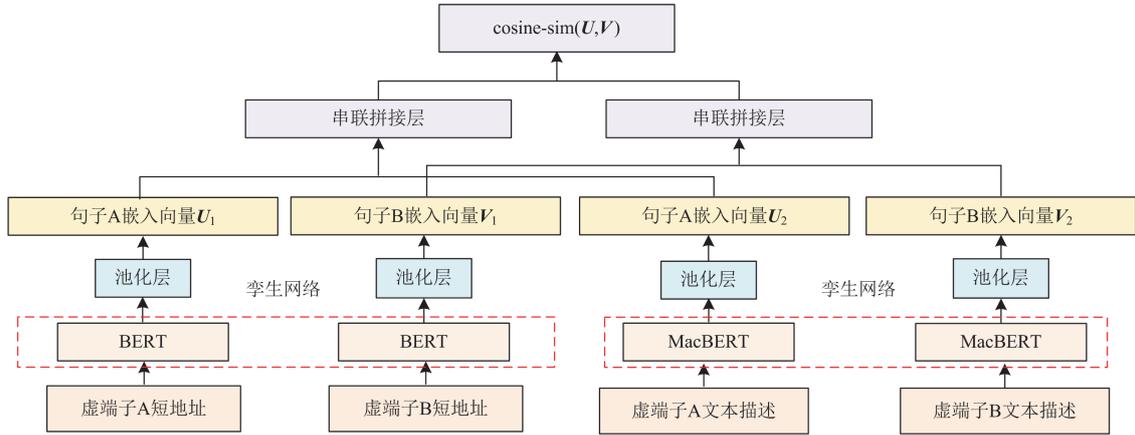


图 4 基于 Sentence-MacBERT 的虚端子自动连接模型

Fig. 4 Sentence-MacBERT-based virtual terminal automatic connection model

### 4.3 虚端子自动连接框架

基于上述 Sentence-MacBERT 模型提出虚端子自动连接技术框架，具体结构如图 5 所示。

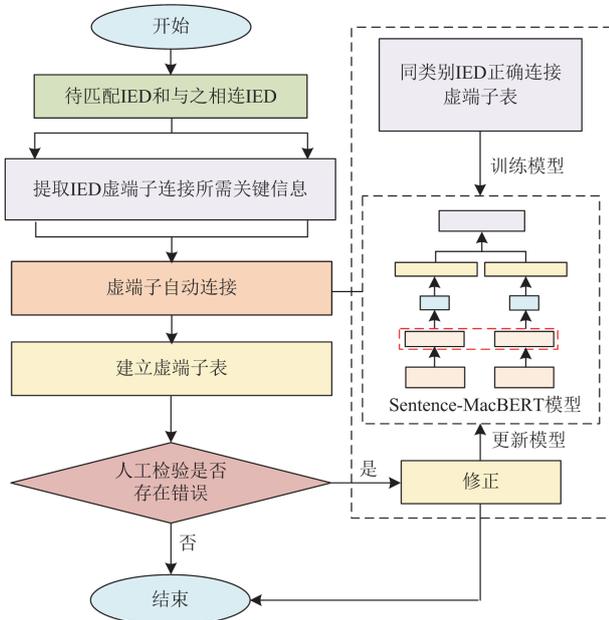


图 5 虚端子自动连接技术框架

Fig. 5 Technical framework for automatic connection of virtual terminals

其具体流程如下：

1) 确定待匹配和与之相连的 IED 设备，分别提

取虚端子自动连接所需关键信息，包括其短地址和中文描述。

2) 利用 MATLAB、Pkuseg 等工具对关键信息进行预处理。

3) 利用同类别 IED 设备已正确连接的虚端子表对 Sentence-MacBERT 模型进行训练，直至达到所设评价指标的评价标准。

4) 将预处理后的数据输入到训练后的 Sentence-MacBERT 模型，实现虚端子自动连接，并生成虚端子表。

5) 检验虚端子表是否存在问题。如不存在，则输出相应虚端子表；如存在，则对虚端子表进行修正后输出，并将修正后的虚端子表重新输入到模型中以更新模型。

## 5 算例分析

### 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 4 所示。

### 5.2 实验数据与评估指标

为验证基于 Sentence-MacBERT 模型的虚端子自动连接方法的有效性，本文以 110 kV 智能变电站

主变测控装置为例, 选取来自不同厂家、不同型号的主变测控装置虚端子短地址与中文描述共 2136 条, 构建数据集进行模型训练。其中正确连接的虚端子样本记为正样本并记为标签 1, 错误连接的虚端子样本记为负样本并记为标签 0, 部分数据集如表 5 所示。

表 4 具体硬件配置

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

表 5 部分虚端子样本数据集

Table 5 Partial virtual terminal sample datasets

输入虚端子		输出虚端子		标签
短地址	中文描述	短地址	中文描述	
SLVD/SVINTCTR1.Amp	高压 1 侧 A 相测量电流	IML1102A.MUSV01/TCTR1.Amp	高压测量电流 A 相 1	1
SLVD/SVINTCTR2.Amp	高压 1 侧 B 相测量电流	IML1102A.MUSV01/TCTR2.Amp	高压测量电流 B 相 1	1
SLVD/SVINTCTR3.Amp	高压 1 侧 C 相测量电流	IML1102A.MUSV01/TCTR3.Amp	高压测量电流 C 相 1	1
SLVD/SVINTVTR1.Vol	高压 1 侧 A 相电压	IML1102A.MUSV01/TVTR1.Vol	高压测量电压 A 相 1	1
SLVD/SVINTVTR2.Vol	高压 1 侧 B 相电压	IML1102A.MUSV01/TVTR2.Vol	高压测量电压 B 相 1	1
SLVD/SVINTVTR3.Vol	高压 1 侧 C 相电压	IML1102A.MUSV01/TVTR3.Vol	高压测量电压 C 相 1	1
SLVD/SVINTVTR10.Vol	低压 1 侧 A 相电压	IML1102A.MUSV01/TVTR10.Vol	低压测量电压 A 相 1	1
SLVD/SVINTVTR11.Vol	低压 1 侧 B 相电压	IML1102A.MUSV01/TVTR11.Vol	低压测量电压 B 相 1	1
SLVD/SVINTVTR12.Vol	低压 1 侧 C 相电压	IML1102A.MUSV01/TCTR12.Vol	低压测量电压 C 相 1	1
SLVD/SVINTCTR10.Amp	低压 1 侧 A 相测量电流	IML1102A.MUSV01/TCTR11.Amp	低压测量电流 B 相 1	0
SLVD/SVINTCTR11.Amp	低压 1 侧 B 相测量电流	IML1102A.MUSV01/TCTR12.Amp	低压测量电流 C 相 1	0
SLVD/SVINTCTR12.Amp	低压 1 侧 C 相测量电流	IML1102A.MUSV01/TCTR10.Amp	低压测量电流 A 相 1	0

为了全面评估模型的性能, 采用准确率  $A$  和  $F_1$  值两项指标对模型进行评价。其中准确率具体计算公式如式(2)所示。

$$A = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (2)$$

式中:  $T_p$ 、 $T_N$ 、 $F_p$ 、 $F_N$  为混淆矩阵元素, 混淆矩阵如表 6 所示。

表 6 混淆矩阵

Table 6 Confusion matrix

预测类别	实际类别	
	1	0
1	$T_p$	$F_p$
0	$F_N$	$T_N$

表 6 中:  $T_p$  代表实际类别和预测类别都为 1 的数据;  $T_N$  代表实际类别和预测类别都为 0 的数据;  $F_p$  代表实际类别为 0 但预测类别为 1 的数据;  $F_N$  代表实际类别为 1 但预测类别为 0 的数据。

然而, 仅仅依靠准确率  $A$  无法全面反映虚端子自动连接模型的识别能力, 因此引入  $F_1$  值作为评估指标,  $F_1$  值是召回率  $R$  和精确率  $P$  的调和均值, 具体计算公式如式(3)一式(5)所示。

$$P = \frac{T_p}{T_p + F_p} \quad (3)$$

$$R = \frac{T_p}{T_p + F_N} \quad (4)$$

$$F_1 = \frac{2PR}{P + R} \quad (5)$$

### 5.3 词向量可视化处理

虚端子中文描述在经过 MacBERT 模型后, 每个单词都被映射为一个高维词向量。然而, 这些高维词向量很难直观地判断 MacBERT 模型生成的词向量的合理性。因此, 本文采用均匀流形逼近与投影算法(uniform manifold approximation and projection, UMAP)对词向量进行降维可视化处理。

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。本文通过 UMAP 算法将高维词向量映射到 3 维空间, 部分结果如图 6 所示。

从图 6 中可以看出, 以“延迟”和“延时”为例, 二者之间的距离很近, 表明二者在语义上是类似的, 这与实际情况相符, 证明 MacBERT 模型生成的词向量包含语义信息且合理。

### 5.4 池化策略选择

Sentence-MacBERT 模型在 MacBERT 和 BERT 模型输出的基础上增加了池化层, 为验证池化策略对实验结果的影响, 本文选择不同的池化策略对 Sentence-MacBERT 模型进行实验, 具体结果如图 7

和图 8 所示。

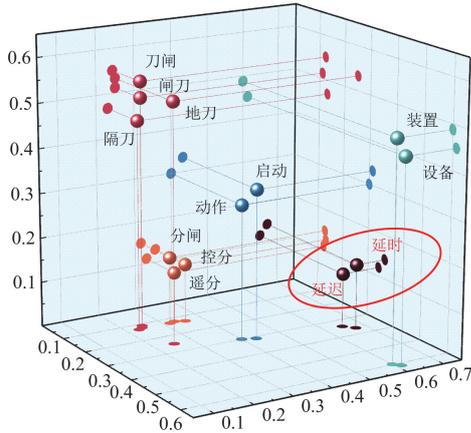


图 6 词向量可视化

Fig. 6 Word vector visualization

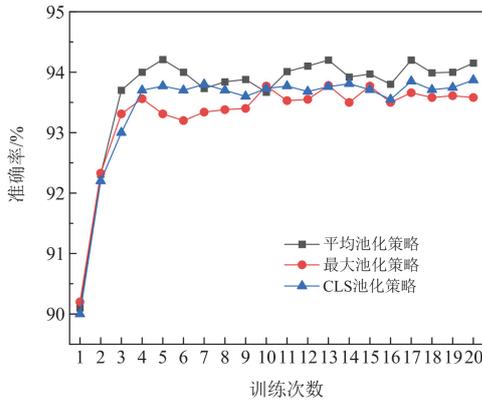


图 7 不同池化策略下的准确率曲线

Fig. 7 Accuracy curves under different pooling strategies

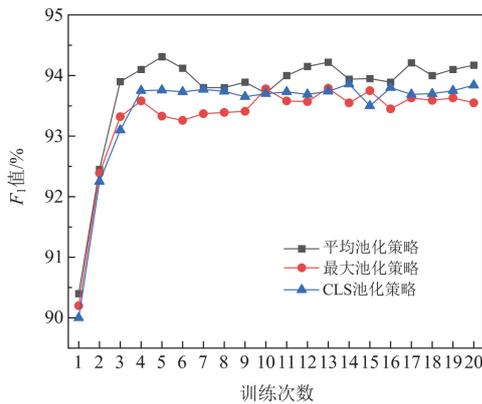


图 8 不同池化策略下的  $F_1$  曲线

Fig. 8  $F_1$  curves under different pooling strategies

由图 7 和图 8 可知，3 种池化策略在初期的几次训练后均达到了较高的准确率和  $F_1$  值。此外，随着训练次数的增加，3 种池化策略均达到相对稳定

的数值。其中，平均池化策略表现最优，在第 17 次训练时准确率和  $F_1$  值均达到了 94.3% 以上，得到了最优模型。因此本文选取平均池化作为池化策略。

### 5.5 模型性能分析

为验证本文提出的 Sentence-MacBERT 模型虚端子自动连接效果，利用上文构建的数据集分别训练 Sentence-BERT 和 Sentence-MacBERT 模型，并对两模型进行不断优化调整。为保证实验的准确性，采用相同的预处理手段对中文描述和短地址进行处理，且池化层均采用平均池化策略。实验参数如表 7 所示。

表 7 Sentence-MacBERT 模型参数

Table 7 Sentence-MacBERT 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
Batch_size	16	批处理尺寸
Dropout_rate	0.1	丢弃率
Max_length	150	MacBERT 最大读取序列长度

使用准确率  $A$  和  $F_1$  值评估模型虚端子自动连接效果，实验结果分别如图 9 和图 10 所示。

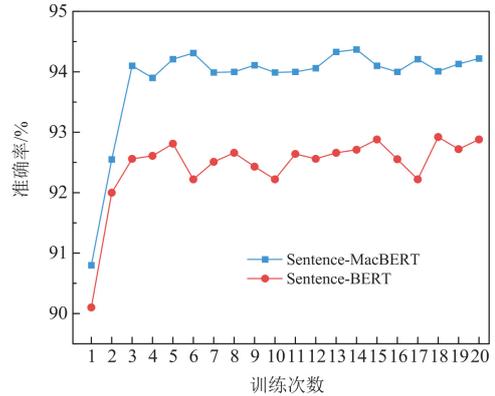


图 9 模型准确率曲线

Fig. 9 Model accuracy curve

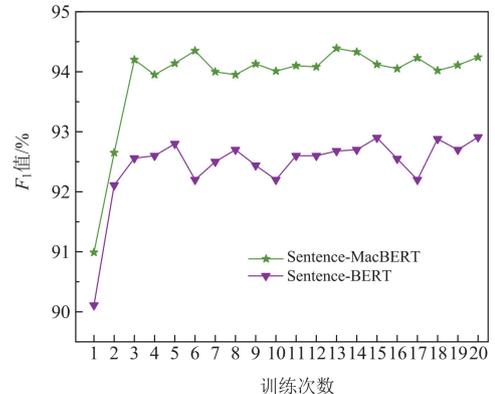


图 10 模型  $F_1$  值曲线

Fig. 10  $F_1$  value curve of the model

从图 9 和图 10 所示的训练过程可见, Sentence-BERT 与 Sentence-MacBERT 在虚端子连接任务上均表现良好。随着训练轮次增加, 两者的准确率与  $F_1$  值持续提升, 约在第 4 轮后进入稳定阶段。其中, Sentence-BERT 在第 18 轮取得最优结果, 准确率与  $F_1$  值分别为 92.92% 和 92.88%; 而本文提出的 Sentence-MacBERT 收敛更快, 在第 14 轮即达到最优, 准确率为 94.37%,  $F_1$  值为 94.33%。对比表明, Sentence-MacBERT 的训练效果显著优于 Sentence-BERT。

为进一步验证 Sentence-MacBERT 的性能优势, 本文将其与若干常用方法进行了对比实验。为保持评估一致性, 所有实验均基于同一主变测控装置构成的测试集, 具体结果见表 8。

表 8 各模型实验结果

Table 8 Experimental results of various models

方法名称	准确率/%	$F_1$ 值/%	所需时间/s
关键字符匹配	86.45	86.46	7.63
改进 Dice 距离	90.23	90.31	3.48
粒子群算法改进距离	91.22	91.31	2.14
烟花算法改进距离	92.74	92.76	2.25
Sentence-MacBERT	94.38	94.41	1.32

从表 8 中可以看出, Sentence-MacBERT 模型相较于其他方法取得了最高的准确率与  $F_1$  值, 分别为 94.38% 和 94.41%。此外, 在实验过程中, Sentence-MacBERT 模型共用时 1.32 s, 相较于其他方法, Sentence-MacBERT 效率更高, 这表明本文所提出的模型具有更好的连接能力, 能够有效地对虚端子进行自动、准确连接。

## 6 结论

本文提出了一种基于 Sentence-MacBERT 模型的虚端子自动连接方法, 通过对虚端子短地址和中文描述进行有效处理, 实现了虚端子的准确连接。

1) 采用了 MacBERT 模型, 并用 UMAP 算法对模型输出结果进行可视化处理, 为后续 Sentence-MacBERT 模型生成句向量提供基础。

2) 构建了基于 Sentence-MacBERT 的虚端子自动连接模型, 并通过仿真选择平均池化作为池化策略。结果表明, 该模型的准确率和  $F_1$  值分别能够达到 94.38% 和 94.41%, 能有效地完成虚端子自动连接任务。

3) 提出了智能变电站虚端子自动连接框架, 帮助设计人员节省了虚端子手动连接所需时间, 提高了智能变电站的设计效率, 具有良好的应用前景。

## 参考文献

[1] 温才权, 韦鑫, 闫茂华, 等. 基于暂态电流概率分析的

- PT 二次多分支电缆中性线接地故障辨识研究[J]. 电网与清洁能源, 2024, 40(12): 45-53.
- WEN Caiquan, WEI Xin, YAN Maohua, et al. Research on neutral line ground fault identification of PT secondary multi-branch cable based on transient current probability analysis[J]. Power System and Clean Energy, 2024, 40(12): 45-53.
- [2] 肖豪龙, 臧欣, 梁文武, 等. 变电站二次电缆绝缘缺陷原因分析与治理措施研究[J]. 湖南电力, 2021, 41(4): 34-37.
- XIAO Haolong, ZANG Xin, LIANG Wenwu, et al. Research on reason and treatment measures of secondary cable insulation defects in transformer substation[J]. Hunan Electric Power, 2021, 41(4): 34-37.
- [3] 赵学奎. 变电站二次电缆的安全性分析及防护措施研究[D]. 沈阳: 沈阳工业大学, 2022.
- ZHAO Xuekui. Safety analysis and protection measures of secondary cables in substations[D]. Shenyang: Shenyang University of Technology, 2022.
- [4] 刘世涛, 徐兆国, 杨凯, 等. 典型缺陷下电力电缆局部放电特征与识别[J]. 高电压技术, 2023, 49(S1): 36-39.
- [5] 许家焰, 武芳瑛, 李瑞, 等. 智能变电站继电保护二次虚回路失效风险动态识别技术[J]. 电力系统保护与控制, 2025, 53(1): 160-170.
- XU Jiayan, WU Fangying, LI Rui, et al. Dynamic identification technology for the failure risk of the secondary virtual circuit of relay protection in an intelligent substation[J]. Power System Protection and Control, 2025, 53(1): 160-170.
- [6] MATHEBULA V C, SAHA A K. Reliability of IEC 61850 based substation communication network architecture considering quality of repairs and common cause failures[J]. Protection and Control of Modern Power Systems, 2022, 7(1): 174-188.
- [7] 俞伊丽, 张展耀, 接晓霞, 等. 基于知识图谱与 SCD 文件的智能变电站二次检修安全措施自动生成技术研究[J]. 电力系统保护与控制, 2024, 52(2): 129-142.
- YU Yili, ZHANG Zhanyao, JIE Xiaoxia, et al. Automatic generation technology of secondary safety measures in an intelligent substation based on a knowledge graph and SCD files[J]. Power System Protection and Control, 2024, 52(2): 129-142.
- [8] 马国明, 王思涵, 秦炜淇, 等. 输电线路运行状态光纤感知研究与展望[J]. 高电压技术, 2022, 48(8): 3032-3047.
- MA Guoming, WANG Sihan, QIN Weiqi, et al. Research and prospect in optical fiber sensing of transmission line operating conditions[J]. High Voltage Engineering, 2022, 48(8): 3032-3047.
- [9] 郑海龙, 吕桂贤, 江覃晴, 等. 基于信息熵理论对变压器光纤传感器声探测信号的特征提取及识别研究[J]. 电力系统保护与控制, 2024, 52(10): 156-166.
- ZHENG Hailong, LÜ Guixian, JIANG Tanqing, et al. Feature extraction and recognition of transformer fiber optic sensor acoustic detection signals based on information entropy theory[J]. Power System Protection and Control,

- 2024, 52(10): 156-166.
- [10] 刘明忠, 童晓阳, 郑永康, 等. 智能变电站配置描述虚端子多视角图形化查看系统[J]. 电力系统自动化, 2015, 39(22): 104-109.  
LIU Mingzhong, TONG Xiaoyang, ZHENG Yongkang, et al. Multi-view graphical viewing system for virtual terminals in smart substation configuration description[J]. Automation of Electric Power Systems, 2015, 39(22): 104-109.
- [11] 古千硕, 丛伟, 王峰, 等. 基于模板复用和特征解析的智能变电站虚回路自动配置方法[J]. 电力自动化设备, 2024, 44(10): 216-224.  
GU Qianshuo, CONG Wei, WANG Feng, et al. Automatic configuration of virtual loop in intelligent substation based on template reuse and feature analysis[J]. Electric Power Automation Equipment, 2024, 44(10): 216-224.
- [12] 高磊, 闫培丽, 阮思焯, 等. 基于相似度计算的学习型模板库在虚回路设计和校验中的应用[J]. 电力自动化设备, 2017, 37(7): 205-212.  
GAO Lei, YAN Peili, RUAN Siye, et al. Application of similarity-calculation-based learning template library in design and check of virtual circuit[J]. Electric Power Automation Equipment, 2017, 37(7): 205-212.
- [13] 李腾, 吕飞鹏, 王星宇, 等. 基于烟花算法计算莱文斯坦距离权重的虚端子自动连接方法[J]. 电测与仪表, 2023, 60(11): 90-94.  
LI Teng, LÜ Feipeng, WANG Xingyu, et al. Virtual terminal automatic connection method for calculating Levenstein distance weight based on fireworks algorithms[J]. Electrical Measurement & Instrumentation, 2023, 60(11): 90-94.
- [14] 武玉萍, 邵雪瑾, 郭科, 等. 基于 LGOS 和 LSVS 逻辑节点的变电站虚端子对接设计研究[J]. 能源与环保, 2021, 43(12): 191-195.  
WU Yuping, SHAO Xuejin, GUO Ke, et al. Research on virtual terminal docking design of substation based on LGOS and LSVS logical node[J]. China Energy and Environmental Protection, 2021, 43(12): 191-195.
- [15] 陈盼, 林传伟, 汤惠芬, 等. 基于关键字符匹配的虚端子自动关联方法[J]. 电力系统保护与控制, 2013, 41(18): 132-137.  
CHEN Pan, LIN Chuanwei, TANG Huifen, et al. A method for virtual terminals automatic connection based on key-character matching[J]. Power System Protection and Control, 2013, 41(18): 132-137.
- [16] 陈桥平, 纪陵, 李一泉, 等. 智能变电站二次设备自动关联技术研究与应用[J]. 广东电力, 2015, 28(12): 74-79.  
CHEN Qiaoping, JI Ling, LI Yiquan, et al. Research and application of automatic association technology for intelligent substation secondary equipment[J]. Guangdong Electric Power, 2015, 28(12): 74-79.
- [17] 王培林, 宋小丽, 王涵宇, 等. 智能变电站虚端子辅助自动设计系统[J]. 电力系统保护与控制, 2020, 48(5): 151-157.  
WANG Peilin, SONG Xiaoli, WANG Hanyu, et al. Aided automatic design system of virtual terminals in intelligent substation[J]. Power System Protection and Control, 2020, 48(5): 151-157.
- [18] 王文琪, 胡炎, 赵娜, 等. 基于距离权重向量优化模型的虚端子自动连接方法[J]. 电网技术, 2018, 42(1): 346-352.  
WANG Wenqi, HU Yan, ZHAO Na, et al. Automatic connection method of virtual terminators based on optimization model of distance weight vectors[J]. Power System Technology, 2018, 42(1): 346-352.
- [19] 戴志辉, 邱小强, 耿宏贤, 等. 基于相似度匹配的智能站虚端子连接准确性判断方法[J]. 华北电力大学学报, 2021, 48(3): 32-38.  
DAI Zhihui, QIU Xiaoqiang, GENG Hongxian, et al. An approach to judge veracity of virtual terminal connection in smart substations based on similarity matching[J]. Journal of North China Electric Power University, 2021, 48(3): 32-38.
- [20] 董怀普, 王位杰, 刘文彪, 等. 基于深度搜索的二次虚实回路融合故障诊断技术[J]. 电力自动化设备, 2019, 39(5): 219-223.  
DONG Huaipu, WANG Weijie, LIU Wenbiao, et al. Fault diagnosis technology of secondary virtual loop and physical fiber loop fusion based on depth first search[J]. Electric Power Automation Equipment, 2019, 39(5): 219-223.
- [21] 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.
- [22] 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.
- [23] 戴志辉, 张富泽, 韩笑, 等. 基于 Sentence-MacBERT 模型的同源录波数据匹配方法[J]. 电力系统保护与控制, 2025, 53(8): 159-167.  
DAI Zhihui, ZHANG Fuze, HAN Xiao, et al. Homologous recording data matching method based on the Sentence-MacBERT model[J]. Power System Protection and Control, 2025, 53(8): 159-167.

收稿日期: 2025-03-31; 修回日期: 2025-07-01

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二次可靠性评估。E-mail: hencepu@163.com

(编辑 魏小丽)