

## 基于三元组网络的新能源监控系统数据智能关联方法

邹花蕾<sup>1</sup>, 曹霞<sup>2</sup>, 陈宇<sup>1</sup>, 葛辉<sup>1</sup>

<sup>1</sup>南京邮电大学自动化学院 江苏南京

<sup>2</sup>南京邮电大学通达学院 江苏扬州

**【摘要】**针对新能源监控系统中信号点关联作业依赖人工配置, 工作强度大、效率低、易出错等实际问题, 本文提出了一种融合深度语义理解与多维特征编码的数据智能关联方法。该方法首先设计了一种基于 BERT 模型的信号名称编码模型, 并创新性地采用三元组网络架构进行对比学习微调, 针对应用场景精准提取语义特征; 设计融合机组编号、信号类型等多维特征的补充编码模型, 并通过基于性能指标的权重分配策略实现模型有效融合, 提升匹配精度。此外, 方法引入“泛型”概念设计了可复用元模板, 进一步提升工程效率。最后, 选择国内三个实际风电场数据进行了实验和测试。结果表明, 本算法使单台风机配置时间缩短 73.4%, 有效减少手动配置点数, 显著提升工程效率, 为新能源场站智能化配置提供有效解决方案。

**【关键词】** 新能源监控; 数据关联; 三元组网络; BERT 模型; 自然语言处理

**【基金项目】** 江苏省自然科学基金项目(BK20230353)

**【收稿日期】** 2026 年 2 月 15 日

**【出刊日期】** 2026 年 3 月 17 日

**【DOI】** 10.12208/j.jeea.20260006

### Intelligent data association method for renewable energy monitoring systems based on triplet networks

Hualei Zou<sup>1</sup>, Xia Cao<sup>2</sup>, Yu Chen<sup>1</sup>, Hui Ge<sup>1</sup>

<sup>1</sup>College of Automation, Nanjing University of Posts and Telecommunication, Nanjing, Jiangsu

<sup>2</sup>Nanjing University of Posts and Telecommunications Tongda College, Yangzhou, Jiangsu

**【Abstract】** To address the practical issues of manual configuration—such as high workload, low efficiency, and error susceptibility—in signal point association operations within renewable energy monitoring systems, this paper proposes an intelligent data association method that integrates deep semantic understanding and multi-dimensional feature encoding. First, a signal name encoding model based on the BERT model is designed, and a triplet network architecture is innovatively employed for contrastive learning and fine-tuning to accurately extract semantic features tailored to specific application scenarios. Additionally, a supplementary encoding model is developed to incorporate multi-dimensional features, such as turbine numbers and signal types, with a performance-based weight allocation strategy to enable effective model fusion and enhance matching accuracy. Furthermore, the method introduces the concept of “generics” to design reusable meta-templates, thereby further improving engineering efficiency. Finally, experiments and tests were conducted using data from three actual wind farms in China. The results demonstrate that this algorithm reduces the configuration time for a single wind turbine by 73.4%, effectively decreases the number of manually configured points, significantly enhances engineering efficiency, and provides an effective solution for the intelligent configuration of renewable energy stations.

**【Keywords】** Renewable energy monitoring; Data association; Triplet network; BERT model; Natural language processing

#### 引言

随着中国能源低碳转型的加速推进, 风力发电和光伏发电领域装机容量出现了爆发式增长。同时, 随着

自动化、信息化等技术在新能源领域的发展, 单个新能源场站往往部署超过 20 个各种功能和应用的系统。一方面, 各系统拥有独立的硬件和软件; 另一方面, 许多

高级应用系统需要从站内的一体化监控系统中获取数据, 以实现自身功能。例如, 在风电场内, 风功率预测系统、风电机组能效分析与故障预警系统、主动支撑控制系统、辅控设备监测系统, 以及各风机设备厂家的信息系统等, 都与风电场的一体化监控系统存在信息交互需求。所以, 根据具体的实际应用情况, 在系统部署和投运阶段, 对应地产生了多种的工程配置任务, 其中就包括了系统间的信号点关联挑点工作。对于新能源场站, 监控平台的信号点接入量巨大, 例如一个 300MW 的光伏电站, 接入点不少于 20 万点, 而同样装机容量的风电场, 一般也有 5 万个的接入点。同时, 各系统间在进行信号点关联的配置时, 同一个信号点在不同系统中的命名并不相同, 目前新能源领域也缺乏输变电领域类似 IEC61850 这样的标准, 可以通过统一的 SCD 文件解析和配置。新能源场站的信号点关联工作往往依赖工程人员手动挑点和配置, 配置工作量巨大, 效率低且易出错, 已成为制约新能源工程项目快速交付与智能化升级的瓶颈之一。

信号点关联的业务数据主要是文本数据, 当需要实现文本数据之间的智能关联时, 可采用自然语言处理 (natural language processing, NLP) 技术搭配机器学习的方法<sup>[1]</sup>。通过文献可知, 目前学界在电力文本数据处理及信息挖掘方面取得了一定的进展<sup>[2]</sup>, 探索了数据挖掘在电力领域的应用。文献[3]提出的基于告警信号文本挖掘的电力调度故障诊断, 文献[4]聚焦于基于深度学习模型的电力设备缺陷文本分类研究, 文献[5-6]研究了基于电力文本、知识图谱的问答系统。文本数据挖掘在电力领域的应用经验为研究信号点数据智能关联方法提供了参考借鉴。另外, 信息检索等领域所采用的文本相似度计算方法<sup>[7-8]</sup>也是本文所提智能关联方法的关键技术。但将相关技术直接应用于解决工业信号点关联这一特定问题仍面临诸多挑战。文献[9]采用 TF-IDF 等词频方法进行关键词匹配, 但难以处理缩写等情况, 以及捕捉实际语义。近年来, 预训练模型如 BERT<sup>[10]</sup>为语义深度理解提供了新范式, 然而, 现有研究<sup>[11]</sup>在工业场景的应用中, 往往仅停留在使用通用 BERT 模型而未进行领域深度适配, 导致模型在实际工程中的准确性和鲁棒性未能达到预期。

为克服上述局限, 本文设计了一种紧贴工程实践、多模型融合的数据智能关联方法。首先, 通过三元组网络对比学习对 BERT 模型进行有监督微调, 使其精准学习电力信号名称的语义空间分布; 其次, 设计多维特征编码模型与语义模型形成互补, 并通过基于性能指

标的权重融合策略, 构建更具鲁棒性的相似度计算框架。最后, 将算法与可复用的元模板设计相结合, 形成了一套从智能推荐到批量复用的完整方法。数据智能关联算法部署于应用系统的配置工具软件, 在工程人员配置过程中, 可减轻手动挑点的工作量, 并降低人工挑点时的错误操作概率, 提升配置准确率。

## 1 关联算法关键技术研究

### 1.1 信号名称编码模型设计

信号名称的相似度判断, 本质是对语义相似度的判断。余弦相似度作为度量指标是有效方式<sup>[12]</sup>, 而获取可用于计算余弦相似度的向量的方法, 本文设计以 BERT 模型为基础的信号名称编码模型。

BERT (Bidirectional Encoder Representations from Transformers) 是一种基于 Transformer 架构的深度双向预训练语言模型。其核心创新在于通过掩码语言模型 (Masked Language Model, MLM) 和下一句预测 (Next Sentence Prediction, NSP) 两个预训练任务, 实现对上下文信息的深度双向理解。与传统单向语言模型不同, BERT 在预训练时同时利用目标词左右两侧的上下文, 从而能够更精准地捕捉语言的细微差别。

BERT 模型的核心运算单元是 Transformer 编码器, 它依赖于自注意力机制。该机制允许模型在处理每个词时, 权衡序列中所有其他词的重要性。所以, 本模型具备自注意力层, 其输出计算如下:

$$Attention(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

其中, Q (查询)、K (键)、V (值) 均是由输入词嵌入通过线性变换得到的矩阵,  $d_k$  是键向量的维度,  $\sqrt{d_k}$  项用于缩放点积, 防止梯度消失。

在预训练阶段, MLM 任务通过随机遮蔽输入序列中一定比例的词汇, 让模型预测被遮蔽的原始词汇, 其目标是 minimized 以下交叉熵损失:

$$L_{MLM} = - \sum_{i \in M} \log P(x_i | x_{\setminus M}) \quad (2)$$

其中, M 代表被遮蔽的词汇索引集合,  $x_i$  是被遮蔽词汇的真实值,  $x_{\setminus M}$  代表未被遮蔽的上下文。

由于输出结果用于信号名称相似度判断这个任务, BERT 的输出量至关重要, 其输出包含序列输出 (每个 Token 的上下文向量) 和池化输出 ([CLS]令牌对应的聚合向量), [CLS]令牌是 Classification Token, 即分类令牌。在通用场景下, 研究者通常使用 [CLS] 向量, 或通过对序列输出进行池化来获得句子表征。在本研究中,

由于拥有结构化的样本数据, 并将采用三元组网络架构对预训练的 BERT 模型进行有监督的微调。基于此微调策略, 在端到端的微调过程中, 三元组损失函数将直接引导[CLS]向量学习到适用于本任务特定分布的高质量、判别性语义表征, 其效能远超静态池化方法。所以, 本文选择 BERT 输出的[CLS]向量作为最终的句子表征。

### 1.2 基于三元组网络的对比学习

为了进一步提升 BERT 模型在语义相似度判断上的判别能力, 实现有效的微调, 本研究采用三元组网络架构对 BERT 模型进行有监督的微调。该方法的核心思想源于对比学习, 旨在通过直接比较正负样本, 将语义空间中的相似关系转化为向量空间中的距离关系<sup>[13]</sup>, 符合本研究的需求。

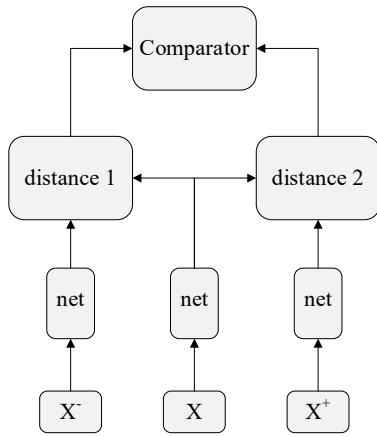


图 1 三元组网络的原理图

三元组网络的结构如图 1 所示, 它接受三个输入: 一个锚点样本、一个与锚点样本语义相似的正样本和一个与锚点样本语义不相似的负样本。这三个样本通过三个共享权重的编码器网络进行前向传播, 分别得到它们的向量表征  $f(A)$ 、 $f(P)$  和  $f(N)$ 。通过三元组网络, 在投影后的向量空间中, 拉近锚点与正样本的距离, 同时增大锚点与负样本的距离。

通过最小化三元组损失 (Triplet Loss) 来实现对正负距离的调整及模型优化。该损失函数的定义如下:

$$L_{triplet} = \max(d(f(A), f(P)) - d(f(A), f(N)) + \alpha, 0) \quad (3)$$

其中,  $d(\cdot, \cdot)$  是一个距离度量函数, 采用欧氏距离或余弦距离。 $\alpha$  是一个超参数, 称为间隔 (margin), 它定义了正负样本对之间应保持的最小距离差。

在本研究中, 编码器  $f(\cdot)$  即为上一节设计的信号名称编码模型, 其输入为监控系统和高级应用系统的待关联信号点名称, 输出为本文选定的 [CLS] 向量。通过将锚点、正例和负例三个句子依次输入 BERT 并

提取其 [CLS] 向量, 再计算三元组损失, 构建了一个端到端的对比学习框架。通过反向传播和梯度下降优化, BERT 模型的所有参数都将被调整, 使得最终生成的输出向量能够精准地编码信号名称的语义信息。

### 1.3 多维特征编码模型设计

针对本场景的特征, 微调后的 BERT 模型, 针对性更强, 实现了通过信号名称的语义相似度比对。而同样具有相似度比对价值的机组编号、信号类型, 没有作为重要信息出现在对比中, 因而本文进一步设计补充模型: 多维特征编码模型, 与前文的编码模型通过权重设置进行模型融合, 得到最终的相似度数值。

多维特征编码模型的设计方式: 选择信号名称、机组编号、信号类型为多维输入特征, 对各个特征以“维度相同、分类处理”的原则设计编码方式, 得到维度相同的向量, 并拼接得到最终的输出矩阵。

具体地, 对信号名称通过去除无意义符号、分词、去除停用词、同义词替换的流程进行特征预处理<sup>[14]</sup>, 得到信号名称对应的词列表 WordList。对于词列表使用词向量方法获取输出量。根据 CBOW (Continuous Bag-of-Words) 模型训练速度快、在小型语料集上表现更好的特点, 选择 CBOW 模型。CBOW 模型是 Word2Vec 框架中的核心架构之一, 旨在通过词的上下文来学习其分布式向量表示。对于信号名称的向量化需求, 一个中心词  $w_t$  及其前后各  $m$  个词构成的上下文窗口 (总大小为  $C=2m$ ), 模型首先将信号名称内所有词的输入向量进行平均聚合, 形成一个统一的上下文语义向量表示:

$$v_{avg} = \frac{1}{C} \sum_{c=1}^C v_{w_c} \quad (4)$$

其中,  $v_{w_c} \in R_d$  表示第  $c$  个上下文词  $w_c$  对应的  $d$  维输入向量,  $v_{avg}$  即为聚合后的上下文表征。该向量随后被送入输出层, 通过计算与词汇表  $V$  中所有词的输出向量的内积, 并经由 Softmax 函数归一化, 得到给定上下文中心词为  $w_t$  的条件概率:

$$p(w_t | Context) = \frac{\exp(v_{w_t}^T \cdot v_{avg})}{\sum_{j=1}^V \exp(v_{w_j}^T \cdot v_{avg})} \quad (5)$$

其中,  $v_w' \in R_d$  是词  $w$  的输出向量, 上标  $T$  表示转置。

本文以电力专业词汇语料库对模型进行训练, 目标是最大化整个语料库中所有此类预测的对数似然。训练后的该模型可以获取信号名称对应的  $m$  维的词向量。

对于数字形式的机组编号, 采用二进制形式表示为向量, 同样取  $m$  维。信号类型采用 one-hot Encoding 方式进行编码, 维度均设置为  $m$ 。如图 2 所示, 由词向量组成的词列表、二进制形式表示的机组编号和编码后的信号类型, 组成了输出矩阵。该矩阵同样采用余弦相似度的方式计算获取语义相似度。

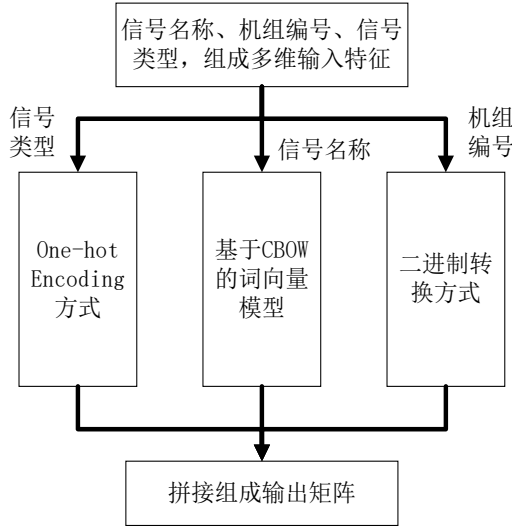


图 2 多维特征编码模型的原理图

### 1.4 可复用的元模板设计

对于本文的新能源场站的工程配置和挑点关联应用场景, 工程人员的配置效率是重要的考量因素, 挑点模板的可复用性设计是工程应用中极具价值的设计思路。本文引入“泛型”模式设计元模板, 完成单个装置的所有信号点关联, 可选择保存并导出挑点模板, 相同装置的挑点工作可复用。

具体地, 信号点关联配置(即选择挑点工作)过程中包括各种业务和逻辑, 比如: 点表排序、标识已选点的处理方法、更新点名、克隆节点的实现、置空点的处理方法、添加节点前的校验等, 按照机组或者装置设置元模板为挑点工作缩减流程。本文在挑点模板的设计中引入计算机科学中的“泛型”概念, 将关键字参数化, 通过在模板中的名称所需位置添加标识符, 标识关键字插入的位置; 并根据实际需求, 设计元模板的其他功能, 包括: 导入导出功能, 用以将挑点模板以 XML 格式文件形式实现导入导出; 顺序记忆功能, 新建的挑点模板可自动记忆挑点序号和顺序。

## 2 整体算法架构介绍

综合前一节所述的算法关键技术, 本文提出的新能源监控系统数据智能关联算法的整体流程如图 3 所示, 其中的信号名称编码模型需要预训练, 它的离线训练流程如图 4 所示。

根据图 3 的流程所示, 在工程配置的过程中, 配置人员选择某个高级应用系统中的某个待配置信号点后, 算法自动检索一体化监控系统中对应装置、对应信号类型的所有可关联信号点, 通过信号名称编码模型, 依次获取所有信号点的含有名称语义信息的输出向量, 并通过余弦相似度依次计算获得待配置信号点名称与所有可关联信号点名称的相似度值。为获取精度更高的信号点相似度值, 本算法中增加了补充编码模型, 对于高级应用系统的待配置信号点和一体化监控系统中所有可关联信号点, 均添加机组编号、信号类型与信号名称共同组成多维输入特征, 通过 1.3 小节的基于自然语言处理技术设计的多维特征编码模型, 获取输出矩阵, 并通过余弦相似度依次计算获得待配置信号点名称与所有可关联信号点名称的相似度值。

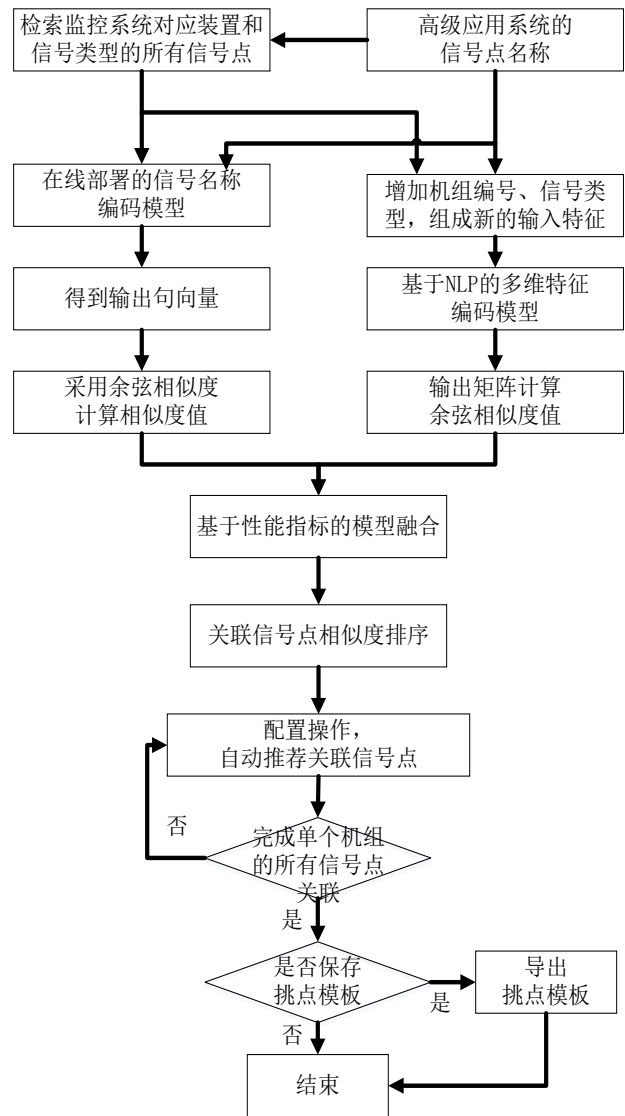


图 3 算法的整体流程图

对于基于两种编码模型获得的相似度值，通过模型融合的方式，获取最终的相似度值。其中，模型融合的方法采用基于性能指标的权重分配法，把准确率（这里的准确率定义为以可关联的信号点中相似度最高的信号点作为输出关联点，与实际关联点对比，在定量数据集上统计正确的输出关联点的比率）作为指标，在一个独立的验证集上评估两个模型的性能，将两个模型的准确率归一化，使其和为 1，从而得到各自权重。

对于模型融合后得到的相似度值，从高到低进行排序，取相似度值最高的两个信号点，输出并展示在配置软件界面的关联配置框的下拉栏中，推荐给工程配置人员，配置人员判断是否采用推荐结果或者手动配置查找正确的关联点。在完成单个机组或装置的所有信号点关联后，可选择保存并导出挑点模板，相同机组或装置的挑点工作可复用。

关联配置为例，风电机组能效分析及故障预警系统需要获取风电场一体化监控系统上的每台风机的历史数据和实时数据，用于风机监视、能效分析、故障分析等功能，因此，在工程配置中，需要将风电机组能效分析及故障预警系统中按风机建模的遥信点、遥测点与风电场一体化监控平台上遥信、遥测表中存储的对应的信号点进行关联匹配。

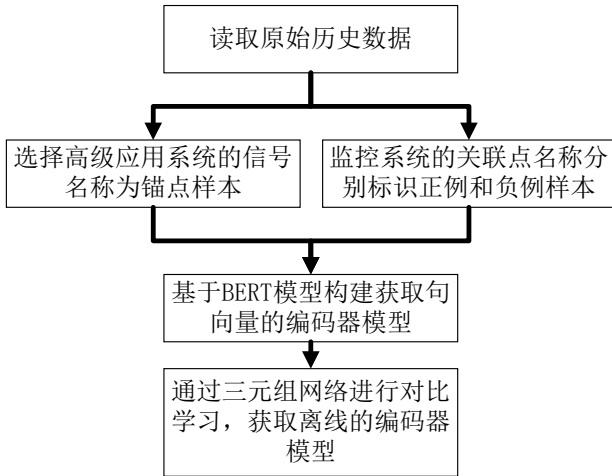


图 4 信号名称编码模型训练流程图

对于图 3 中的信号名称编码模型，它的预训练流程如图 4 所示，采用基于三元组网络的对比学习方法，构建信号名称编码模型预训练流程。首先，从已完成正确配置的原始历史数据库中读取相关信号点名称，并选择高级应用系统的信号名称作为锚点样本。随后，根据监控系统的关联点名称，分别标识正例样本与负例样本，构建用于对比学习的训练数据。在此基础上，基于 BERT 模型架构初始化编码器，通过三元组损失函数优化模型参数，使其能够有效区分不同语义特征的样本，最终训练完成得到信号名称编码模型，为相似度判断任务提供高质量的句向量输出。

### 3 仿真实验测试

#### 3.1 测试环境及数据

本文以风电场高级应用系统“风电机组能效分析及故障预警系统”与风电场一体化监控系统的信号点

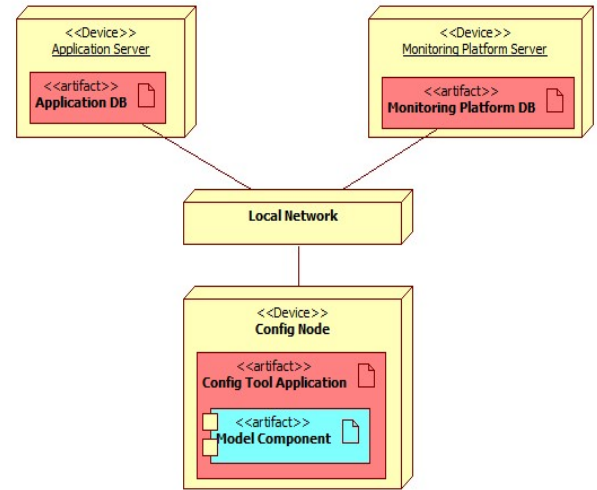


图 5 测试环境的 UML 部署图

按照图 5 的 UML 部署图进行测试环境搭建，模拟实际工程应用的工作环境。在局域网中，准备两台服务器，一台笔记本电脑，一台交换机。两台服务器（图中的 Application Server、Monitoring Platform Server）分别部署风电机组能效分析及故障预警软件系统和风电场一体化监控软件系统，两个系统的数据库均部署在各自的服务器上。将实现了本文算法的动态库或者加密后的脚本程序，部署于风电机组能效分析和故障预警系统的配置工具软件中，并将配置工具软件部署于笔记本电脑上（图中的 Config Node）。

本文选取国内某发电集团所属的 3 个风电场站的历史数据进行测试，具体包括一个陆上风电场（装机容量 60MW）记为 1 号风电场、一个近海风电场（装机容量 300MW）记为 2 号风电场，一个深海风电场（装机容量 500MW）记为 3 号风电场。

#### 3.2 评价指标

本研究的目的，是为了减少配置人员在现场调试工作中的实际工作时间和强度，提高配置效率，从而节约工程成本。因此，首先选取单场站的关联点手动配置减少数量 C1 为评价指标，衡量一个场站的实际工程配置工作中的直接效果。

其次, 为了更科学、客观地评价和比较不同模型的关联点推荐效果, 使用单台风机的配置时间缩短比率  $R_T$  为评价指标:

$$R_T = (1 - \frac{T_N}{T_O}) \times 100\% \quad (6)$$

其中,  $T_O$  是单台风机全部手动配置关联点的总时间,  $T_N$  是使用智能关联算法后的关联点配置的总时间。

### 3.3 测试实验

#### 实验一

实验一使用 3.1 节中所述的三个风电场的数据通过本文算法进行关联点智能推荐的测试。根据风机厂家和型号的不同, 需配置的信号点情况如下: 1 号风电场 3MW 的风电机组共 20 台, 每台需配置的关联点数 420 个信号点。2 号风电场 7MW 的风电机组 22 台, 6.2MW 的风电机组 24 台, 每台需配置的关联点数均为 500 个信号点。3 号风电场 8.3MW 的风电机组 30 台, 每台需配置的关联点数均为 470 个信号点, 6.8MW 的风电机组 37 台, 每台需配置的关联点数为 500 个信号点。三个场站的数据通过部署在配置工具中的本文算法进行关联点推荐, 并与实际的关联点比较, 推荐信号点是正确的数量, 即关联点手动配置减少数量  $C1$  的统计情况详见下表 1。

表 1 实验一的结果统计表

试验数据	需配置总点数/(个)	手动配置减少数量/(个)
1 号风电场	8400	5964
2 号风电场	23000	16155
3 号风电场	32600	22493

#### 实验二

实验二进行不同关联点预测算法的效果比较测试。对比方法分别选择词袋模型 (Bag of Words)、TF-IDF 算法 (Term Frequency-Inverse Document Frequency)、Word2Vec 方法、BERT 模型, 依次生成含有语义的输出向量, 再通过余弦相似度计算语句相似度, 取相似值最高的为推荐结果, 使用单台风机的配置时间缩短比率  $R_T$  为评价指标, 与本文算法进行对比。根据所选评价指标, 单台风机的配置点数需相同才能实现有效对比。由于 2 号风电场的两种类型风机均为 500 个关联点, 因此实验二选用 2 号风电场的数据, 测试结果如表 2 所示。

表 2 实验二的结果统计表

算法名称	$R_T$ 数值/(%)
本文算法	73.4
Bag of Words	40.8
TF-IDF	49.1
Word2Vec	55.3
BERT	64.1

### 3.4 结果分析

首先, 从实验二的结果可以看出, 本文模型的单台风机的配置时间缩短比率最高, 即关联信号点的推荐准确率最高。BERT 模型的配置时间缩短比率排在第二, 而单纯基于词频的方法 (如 Bag of Words 和 TF-IDF), 在本场景下均效果不佳。

第二, 本应用场景是一种对模型准确率要求不高即可产生显著应用效果的非对称场景模式, 如实验 2 中模型推荐准确率最高的本文算法, 也仅为 73% 左右, 但同时单台风机的配置时间缩短也相应缩短了 73%, 即工程配置人员每配置 100 个信号点, 只需要手动查找选择 27 个点, 节省了 73% 的工作时间。结合实验二的结果可以看出, 对于装机容量越大的场站, 该方法对工程配置人员的帮助效果就越明显: 500MW 的 3 号风电场, 针对信号点关联这项工作, 工程人员可以减少 22493 个信号点的手动查找操作。即使是 60MW 的 1 号风电场, 工程人员也可以节省将近 6000 个点的手动配置操作时间, 工程应用效果显著。

### 4 结束语

本研究针对新能源监控系统信号点关联配置的工程效率瓶颈, 提出并实现了一种基于三元组网络的新能源监控系统数据智能关联方法。通过构建基于 BERT 与三元组网络的信号名称编码模型, 并结合机组编号、信号类型等结构化特征构建了补充模型, 算法在复杂工业命名场景下实现了高精度的语义相似度匹配。仿真实验与测试结果表明, 在本应用场景这种对模型准确率要求不高即可产生显著应用效果的非对称场景模式下, 本算法确实显著降低了人工配置工作量, 而且对于装机容量越大的场站, 该方法对工程配置人员的帮助效果就越明显。同时, 本研究仍存在进一步探索的空间, 例如, 在微调阶段融入电力知识图谱以深化模型的领域认知。综上所述, 本研究为工业自动化配置的智能化升级提供了一条行之有效的技术路径, 具有良好的推广应用前景。

## 参考文献

- [1] 王慧芳, 曹靖, 罗麟. 电力文本数据挖掘现状及挑战[J]. 浙江电力, 2019, 38(3): 1-7.
- [2] 邱剑. 电力中文文本数据挖掘技术及其在可靠性中的应用研究[D]. 杭州: 浙江大学, 2016.
- [3] 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.
- [4] 李明, 张伟. 基于 BERT 和 CNN 的电力设备缺陷文本分类方法[J]. 电力系统自动化, 2020, 44(10): 156-163.
- [5] 贾宗泽. 基于知识图谱的电力知识问答与实现技术[D]. 北京: 华北电力大学, 2024.
- [6] 刘沿娟, 张栋栋, 于海亮, 等. 基于知识图谱的电力标准智能问答系统研究[J]. 电工技术, 2024(16):143-146.
- [7] Li, J., Wang, Y., & Sun, H. A Survey on Natural Language Processing for Industrial Systems[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108-125.
- [8] Yujia Wu, Jun Wan. A survey of text classification based on pre-trained language model[J]. Neurocomputing, 2025, 616(1): 128921.
- [9] Li, G., & Sun, C. Bag-of-Words vs. Graph vs. Sequence vs. Convolutional Models for Text Classification[J]. ACM Transactions on Intelligent Systems and Technology, 2023, 14(3): 1-25.
- [10] DEVLIN J, CHANG M W, 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: Human Language Technologies. 2019: 4171-4186.
- [11] CHEN X, LI S, ZHOU F. Intelligent Recognition of Power Customer Complaint Text Using a BERT-CRF Model[J]. Applied Energy, 2022, 318: 119212.
- [12] MINAEI, S., KALCHBRENNER, N., CAMBRIA, E., et al. From TF-IDF to BERT: A Technical Overview of Text Representation Methods in NLP[J]. IEEE Transactions on Artificial Intelligence, 2022, 3(2): 127-145.
- [13] HOFFER, E., & AILON, N. Deep metric learning using triplet network[C]// Proceedings of the 3rd International Conference on Learning Representations (ICLR). 2015.
- [14] 吴桂龙, 杨志敏, 黄昱. 电力通信运行管理中典型业务数据的智能关联方法[J]. 电信科学, 2021, 2: 164-172.

版权声明: ©2026 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS