电力通信故障调度语音识别及语义记录研究

苏云霞[1] 游兆阳[1] 邢 进[1] 任赛男[1] 宋长平[1]
SU Yunxia YOU Zhaoyang XING Jin REN Sainan SONG Changping

摘 要 电力通信缆线故障勘测与应急调度语音通常伴随环境噪声、口音差异和大量专业术语，传统人工记录效率低、易遗漏，而通用语音识别方法无法适用于电力场景使用需求。因此，文章提出一种面向电力通信调度的语音识别及语义记录方法。首先构建包含多噪声场景与方言特征的通信调度语音数据集，并通过数据增强提升语料丰富性。再训练电力通信调度专用语音识别模型，实现对调度语音的识别；然后构建包含真实噪声扰动的文本数据集，结合统计语言模型与序列生成式模型设计文本纠错与语义补全机制，恢复被噪声影响的关键术语和结构化信息。实验结果表明，所提方法可显著提升通信缆线故障调度语音记录转文本的准确率，减少专业表达识别偏差，为调度过程留痕、现场作业复核、应急处理效率提升提供技术支撑。

关键词 电力通信缆线勘测；语音识别；语义融合；深度学习；*N*-gram 模型

doi：10.3969/j.issn.1672-9528.2026.07.015

0 引言

电力通信系统是电网安全稳定运行的核心保障。作为信息传输的主要载体，电力通信缆线的安全直接影响电力通信网络的规划与运维质量[1]。在线路故障勘测与应急调度过程中，现场与调度员通过语音传递勘测指令、反馈现场工况等关键信息，因此，调度语音的准确识别是实现调度自动化、提升电力通信作业效率的重要前提。

传统电力通信调度语音整理主要依靠人工，时间成本高且准确率无法衡量。现有自动语音识别（ASR）模型大多适用于通用场景，无法满足电力通信线路勘测调度场景中大量专用词汇（如“断路器”“负荷”“光缆型号”等）与缩略语的识别需求。

本文聚焦电力通信线路语音调度场景，提出一种端到端深度学习语音识别与语义融合方法。通过构建专用语料库训练基础语音识别模型，引入语义纠错模型优化方法，提升电力调度语音识别的准确性与效率，为电力通信调度自动化升级提供了有效技术支撑，对保障智能电网通信系统稳定运行、推动电力行业数字化转型具有一定的理论与实践意义。

现阶段随着深度学习技术的不断发展，基于卷积神经网络的语音识别技术已在大量通用场景中取得显著成果，语音识别的有效性得到验证。其中 Wav2Vec2.0 模型[2]能够基于自监督学习从原始语音信号中提取特征，大幅提升了语音识别的性能，是当前 ASR 领域的主流模型之一；另一方面，BART 模型基于双向自回归 Transformer 架构[3]设计，具备智能文本生成与自纠错能力，在自然语言处理的众多任务中同样展现出优异性能。

因此，已有大量面向电力通信调度的研究成果涌现。例如在调度通信架构与路由策略研究方面，李莉等[4]针对新型电力系统业务的多样化需求，提出了基于次梯度超平面映射的电力双模路由调度方法，构建了最大化业务映射回报率的双模调度模型，用于解决传统电路交换机制无法满足多业务接入的难题；许兵兵[5]则利用油田电网光纤通信资源，构建了基于 IP 电话技术油田区域电力调度语音融合通信系统，实现了调度通信的网络化管理与标准化运维。

此外，在电力行业语音识别领域，现有成果已形成多元技术体系。例如，张聚明等[1]融合 5G 核心网、智能语音（ASR&TTS）及 5G 音视频调度技术，设计了电力调度融合通信系统，实现有线与 5G 网络的无缝对接，从而提升调度的多媒体通信能力；蔡磊晓等[6]提出将软交换技术、VoIP 网关技术与远端模块技术应用于智能电网语音交互，且构建了电网调控专业语料库，用于实现语音生成操作指令、语音下令回令等功能；王邦鸿[7]则针对国网电力调度场景的软硬件限制与数据保密性需求，设计了实时流多级语音预处理算法，降低了系统解码功耗与实时率，提升了终端侧语音识别的鲁棒性；胡州明等[8]基于动态时间规整算法与自然语言处理技

1. 国网上海市电力公司信息通信公司 上海 200072
[基金项目] 国网上海市电力公司科技项目（B3090F24001K）

术，提出了新能源电站调度语音识别方法，通过词袋特征提取与多标签分类，实现上下文关联分析，提升了方言与专业术语的识别准确率。为了进一步提升调度语音语义准确性，苏云霞等[9]提出基于语义引导的电力通信调度通话语音摘要生成模型，通过语音关键段切分、Wav2Vec2.0 模型微调与 TextRank 语义分析，实现专业通话内容的摘要生成，为调度日志自动化提供了技术支撑。

虽然已有研究在电力调度语音识别、通信架构优化等多个方面均取得进展，但尚未形成适合电力通信线路调度场景的专用语音识别与语义融合方案，尤其在面向专业语料库的复杂噪声环境下的识别鲁棒性、语音语义与调度业务的深度融合方面，仍需进一步研究。

本文提出一种语音识别与语义融合的两级模型架构，包含电力通信调度的 ASR 基础模型与电力通信调度文本纠错模型。算法流程如图 1 所示，通过语料库构建、基础识别、文本纠错、语义优化的逐级处理，实现从原始调度语音到最终文本的端到端转换。具体设计包括带噪语料库构建、ASR 基础模型训练、带噪文本数据集构建、*N*-gram 统计模型建立及文本纠错模型优化五个主要步骤，各步骤共同实现专业语音的识别与语义融合。

图 1 本文语音识别方法实施流程

1 带噪电力通信调度语料库和文本数据集构建

语料库的完备性和文本数据集的可靠性是模型训练的基础。为确保模型能有效应用于电力通信缆线勘测调度实际场景，本研究分别构建了包含真实场景特征的带噪电力通信调度语料库和专业文本数据集。

带噪电力通信调度语料库的构建包含多个步骤，首先收集多源语音数据。收集的语音数据样本覆盖电力通信缆线勘测调度的典型场景，包含地方口音、现场噪声（如施工噪声、设备运行噪声）和语音高低差异等真实特征。之后所有样本经专业语音校对人员进行撰写标注，并通过质量检验，确保标注的准确性，初步构建电力通信调度语音数据库。初步得到的数据库样本具备三个主要特征：一是包含大部分电力通信缆线勘测相关业务名称的明确表述（如“光缆熔接”“缆线路由勘测”等）；二是涵盖特定区域内变电站、光缆型号、勘测设备、电杆等固定命名方式；三是包含不同地区对相同名称、数字的读音差异，确保样本的场景多样性。

针对已收集样本需实施数据增强操作，通过速度扰动、噪声添加、音调调整以及频谱扩充四种方式对语音数据库样本进行数据增强：例如，将原语速样本采用 0.9~1.1 倍变速随机调整，以模拟不同调度人员的语速差异；对原样本加入高斯噪声、椒盐噪声，实现现场复杂噪声环境的模拟；或将原声音的音调高低进行 0.5~1.5 倍间随机调整，模拟不同发音人的声线特征；以及对原样本进行频谱扩充，丰富语音特征维度。通过上述增强操作，数据库中样本数量被扩充为原来的 3 倍，显著提升了样本多样性。

最后，在已有原声语音的基础上补充合成语音样本。基于已有真实电力通信调度语料数据，通过语音合成技术随机生成合成语音样本并加入数据库，能够进一步扩大语料库的规模，保证模型训练的充分性。

另一方面，本文还构建了带噪电力通信调度文本数据集，增强 ASR 基础模型输出文本的纠错能力。数据集以电力通信调度场景下的错误文本样本为输入，正确文本样本为标签，用于文本纠错模型的对抗训练。错误样本及其对应标签采用两种方式获取。

第一种方式通过提取 ASR 基础模型的错误识别文本获取。将 Wav2Vec2.0 模型输出的错误语音识别文本样本与对应的正确文本进行组对，建立错误样本－标签对。例如，若语音被识别为“当前符合曲线分析，预计负载波动较大。”，其正确标签为“当前负荷曲线分析，预计负载波动较大。”，二者组成一组样本对；再如，针对缆线勘测场景，若识别文本为“光缆短路易出现故障”，正确标签为“光缆断路易出现故障”，构建对应的样本对。

第二种方式采用人为生成常见专业转换错误样本。结合电力通信缆线调度的专业术语特点，人为设计包含典型错误的语音短句，重点覆盖易混淆的专业词汇转换错误。例如，将“fuhe”（“负荷”）转换为“符合”“duanluqi”（“断路器”）转换为“短路器”“guanglan duanlie”（“光缆断裂”）写成“光缆短裂”等，确保数据集能够覆盖场景内的主要识别误差类型。

2 基于 Wav2Vec2.0 的 ASR 基础模型训练

完成训练语料库和文本数据库构建后，需要对语音识别基础模型进行训练。模型训练和推理流程如图 2 所示。

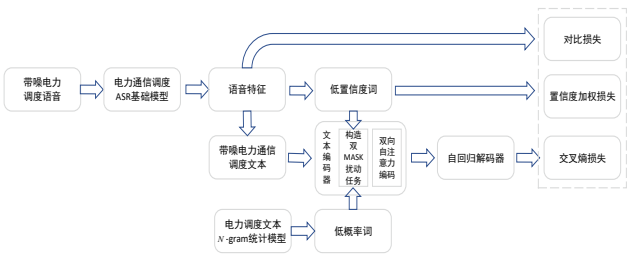

图 2 本文模型的训练和推理流程示意图

其中，因 Wav2Vec2.0 模型的预训练模型在通用语音识别任务中表现优异，因此本文基于该模型训练适配电力通信

缆线勘测调度场景的 ASR 基础模型。首先以自建的带噪电力通信调度语料库为训练数据，输入电力通信调度语音模型，对应的文本标注作为输出标签，冻结 Wav2Vec2.0 Large 部分网络结构，实现模型微调。通过训练，模型可学习电力通信专用语音特征与文本之间的映射关系，实现从原始调度语音到文本的初步转换，同时避免了过拟合，从而得到电力通信调度 ASR 基础模型，为后续文本纠错环节提供输入数据。

3 电力调度文本 *N*-gram 统计模型建立

为使模型进一步适用于电力通信专用词汇的语义理解，本文采用 *N*-gram 统计模型分析电力通信调度文本，构建电力调度文本的 *N*-gram 统计模型，通过计算常用电力通信调度词序列的出现概率，为文本纠错提供语义约束。

N-gram 统计模型纠错文本来源于历史电力通信缆线勘测调度中心通话转录文本、调度文本记录、勘测事故报告及处理文档等，保证了数据的专业性与场景适配性。在此基础上对 *N*-gram 统计模型进行预处理，采用 jieba 分词工具对收集的文本资料进行数据清洗与预处理，包括将连续文本拆分为独立词汇，去除无关字符，如特殊符号、冗余空格等，以及同义词归一化等。例如将“缆线路由勘测”与“光缆路由勘测”统一为标准表述，确保数据的规范性。

考虑到上下文关联性分析的有效性，本研究选择 2-gram 和 3-gram 组合模型，计算电力通信调度专业词汇的共现概率。例如，在 2-gram 模型中计算“负荷－调度”“光缆－断裂”等词序列的共现概率，在 3-gram 模型中计算“缆线路由－勘测－方案”等长词序列的共现概率。

4 基于 BART 架构的文本纠错模型构建与优化

最后，以 BART（bidirectional and auto-regressive transformers）模型架构为基础，结合上述构建的带噪文本数据集与 *N*-gram 统计模型，构建电力通信调度文本纠错模型，并通过对抗训练、双 MASK 扰动策略与多元加权目标函数实现模型优化。

模型训练流程主要包括输入编码、MASK 扰动与文本生成 3 个环节。先将带噪电力通信调度文本输入 BART 模型的文本编码器，再通过双 MASK 扰动策略对输入文本进行 MASK 处理。最后，将处理后的文本输入自回归解码器进行精准文本生成。

此外，本文还提出双 MASK 扰动策略，结合 ASR 基础模型的置信概率与 *N*-gram 统计模型的词序列概率设置 MASK 位置，以提升模型对低置信度文本与语义不合理文本的纠错能力。一方面，设定置信度阈值 c_thr，若 ASR 基础模型输出文本中某词汇的置信概率低于 c_thr，则判定为低置信度词，对其进行 MASK 处理；另一方面，将 *N*-gram 结果用于双 MASK 策略中的第二重 MASK，当某词序列共现概率低于 n_thr 时，将其标记为潜在错误，认为是语义不合理词汇，并对其进一步进行 MASK 处理。通过上述双重 MASK 约束，引导模型重点关注易错词汇与语义异常词汇，提升纠错针对性。

本研究的总损失函数由交叉熵损失、置信度加权损失和对比损失三部分组成，从而保证文本恢复能力、低置信度词修正能力与语义流畅性：

$$L = \alpha L_{\text{ce}} + \beta L_{\text{conf}} + \gamma L_{\text{contrast}} \tag{1}$$

式中：α、β、γ 分别为各损失项的权重系数，可通过实验调试确定最优值。其中，交叉熵损失 L_{ce} 为 BART 模型的常规损失函数，用于衡量模型预测文本与真实标签之间的差异，确保模型具备基本的文本恢复能力。置信度加权损失 L_{conf} 用于强化模型对低置信度词的修正能力，使模型输出偏向于电力通信调度专业用语。

$$L_{\text{conf}} = -\sum_t c_t \cdot y_t \log \hat{y}_t \tag{2}$$

式中：c_t 为 ASR 基础模型计算的第 t 个文本词汇的置信度，y_t 为第 t 个词汇的真实标签值，$\hat{y}_t$ 为模型对第 t 个词汇的预测值。对比损失 L_{contrast} 用于优化文本语义特征，提升输出文本的语义流畅性。

$$L_{\text{contrast}} = -\log \frac{e^{\cos(\boldsymbol{h}_i, \boldsymbol{h}_j)}}{\sum_k e^{\cos(\boldsymbol{h}_i, \boldsymbol{h}_k)}} \tag{3}$$

式中：$\boldsymbol{h}_i$ 是 ASR 基础模型输出的语音特征，$\boldsymbol{h}_j$ 是文本纠错模型输出的文本特征，$\boldsymbol{h}_k$ 是其他样本的文本特征，cos(·) 是余弦相似度计算。通过对比损失约束，使正确文本特征与对应语音特征的相似度高于其他样本，提升语义融合效果。

5 实验数据与分析

本文采用了两类音频数据开展实验验证，包括公开普通话数据集 AISHELL-1 和自建电力通信调度语料库，自建电力通信调度语料库构成如表 1 所示。数据库样本环境噪声采用 THCHS-30 数据集中自带的咖啡馆场景噪音，所有音频均为单声道 wav 格式，采样频率均为 16 kHz。本文实验基于主频 2.20 GHz 的 Intel(R) Xeon(R) Silver 4210 处理器和 NVIDIA Tesla A100 图形处理器；软件环境基于 Linux 操作系统构建，深度学习模型相关代码通过 PyTorch 框架实现。

表 1 自建语料库详细构成

语料库	音频数量 / 条	音频时长 /h
训练集	3 000	24.0
验证集	600	5.4
测试集	1 240	9.8
总计	4 840	39.2

本文模型与U-Conformer、Whisper、Paraformer三个主流声学模型在AISHELL-1公开普通话数据集及自建电力通信调度语料库上的错字率对比结果如表2所示。首先第一列在AISHELL-1数据集上，各模型整体错字率均很低，其中Paraformer模型以3.4的错字率位列第一，体现了其对通用普通话语音的识别能力。本文模型错字率为4.6，虽略高于Paraformer，但优于U-Conformer（5.0）和Whisper（5.6），说明本文模型在通用语音识别任务中也具备较强的竞争力。

相比公开普通话数据集，本文提出的方法在自建的电力调度语料库上优势更为明显。如表2所示所有模型的错字率相比通用普通话数据集均有明显上升，因为电力调度场景不仅专业术语密集、语音指令简洁，而且背景环境复杂，增加了识别难度。尽管如此，本文模型仍以15.8的错字率取得最优表现，较主流Paraformer模型相对降低了3.7%的CER。这一结果表明，本文模型通过有针对性的模型设计与训练优化，更适用于电力通信调度这一特定专业场景的语音识别需求，对专业领域的语音特征捕捉和语义理解能力更具优势。

表2 本文与主流声学模型的错字率（CER）对比结果

模型	AISHELL-1/%	自建语料库/%
U-Conformer	5.0	17.6
Whisper	5.6	17.2
Paraformer	3.4	16.4
本文模型	4.6	15.8

6 结语

本文针对电力通信调度场景中语音识别准确率低、通用模型适配性差的问题，提出了一种端到端基于深度学习的语音识别与语义融合方法。通过构建带噪专用语料库与文本数据集，实现Wav2Vec2.0模型语音识别基础模型的构建，并基于BART架构与*N*-gram统计模型进行文本纠错，提出双MASK扰动策略与多元加权目标函数优化，提升模型自身纠错能力，为电力通信调度智能化提供了可靠的技术方案。

未来研究可进一步拓展语料库的覆盖范围，包含更多极端场景的语音样本，如方言等；并优化模型架构，探索Transformer变体模型在本场景的应用；结合实时语音处理技术，实现调度语音的实时识别与纠错，进一步提升技术的工程应用价值。

参考文献：

[1] 张聚明，杨光，张延斌，等．基于5G网络和智能语音的电力调度融合通信系统研究[J]. 数字通信世界，2024(12):90-92.

[2] Lewis M, Liu Yinhan, Goyal N, et al. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.Brussels:ACL,2020:7871-7880.

[3] Baevski A, Zhou H, Mohamed A, et al. Wav2Vec 2.0: a framework for self-supervised learning of speech representations[PP/OL].(2020-10-22)[2026-06-22].https://doi.org/10.48550/arXiv.2006.11477.

[4] 李莉，孙海波，陆珊珊，等．基于次梯度超平面映射的电力双模路由调度策略[J]. 内蒙古电力技术，2025,43(5):30-36.

[5] 许兵兵．油田区域电力调度语音融合通信系统[J]. 上海电气技术，2024,17(3):18-21.

[6] 蔡磊晓，张杰．语音交互技术在智能电网中的应用研究[J]. 电声技术，2023,47(12):40-42.

[7] 王邦鸿．基于国网电力调度场景下的智能语音识别系统的实现与优化[D]. 西安：西安电子科技大学，2024.

[8] 胡州明，唐冬来，李玉，等．基于自然语言处理的电力调度语音识别方法[J]. 微型电脑应用，2023,39(6):171-174.

[9] 苏云霞，游兆阳，邢进，等．基于语义引导的电力通信调度通话语音摘要生成[J]. 电力信息与通信技术，2025,23(10): 1-7.

【作者简介】

苏云霞（1991—），女，河南商丘人，硕士，工程师，研究方向：电力通信网络维护、调控数据分析和管理。

（收稿日期：2025-12-05 修回日期：2026-07-10）