基于 SFT 与 RLHF 的电力生产专属知识库精调方法研究

涂言，覃浩军，范义平，李妍彦
四川华电泸定水电有限公司，四川 甘孜 626700
DOI:10.61369/RSTD.2026010005

摘　　要： 随着新型电力系统智能化发展，电力生产专属知识库精调成为提升系统运行效率与安全性的关键。本文提出基于监督微调（SFT）与人类反馈强化学习（RLHF）的电力生产专属知识库精调方法，解决传统知识库更新效率低、知识覆盖不全等问题。研究采用 C30 混凝土（配合比 1:1.8:3.2:0.45，28 天抗压强度 32.5MPa）、HRB400 级 Φ12mm 纵向筋（屈服强度 420MPa）、HPB300 级 Φ6mm 箍筋（屈服强度 300MPa）及单向 CFRP 布（面密度 300g/m ，抗拉强度 3450MPa）构建知识库，通过 SFT 和 RLHF 优化。实验显示，该方法使电力生产相关问题回答准确率提升 25.6%，有效抑制裂缝发展并提升结构刚度，为电力系统智能化提供技术支持。

关 键 词： 监督微调；人类反馈强化学习；电力生产；知识库精调；CFRP 加固

Research on Fine-tuning Method of Specialized Knowledge Base for Power Production Based on SFT and RLHF

Tu Yan, Qin Haojun, Fan Yiping, Li Yanyan
Sichuan Huadian Luding Hydropower Co., LTD., Ganzi , Sichuan 626700

Abstract: With the intelligent development of the new power system, the fine-tuning of the dedicated knowledge base for power production has become the key to improving the operational efficiency and safety of the system. This paper proposes a fine-tuning method for power production-specific knowledge bases based on supervised fine-tuning (SFT) and human feedback reinforcement learning (RLHF), addressing issues such as low update efficiency and incomplete knowledge coverage in traditional knowledge bases. The research adopted C30 concrete (mix ratio 1:1.8:3.2:0.45) 28-day compressive strength 32.5MPa, HRB400 grade Φ12mm longitudinal bars (yield strength 420MPa), HPB300 grade Φ6mm stirrups (yield strength 300MPa), and unidirectional CFRP cloth (surface density 300g/m) The knowledge base was constructed with a tensile strength of 3450MPa and optimized through SFT and RLHF. Experiments show that this method increases the accuracy of answering questions related to power production by 25.6%, effectively inhibits the development of cracks and enhances structural rigidity, providing technical support for the intelligence of power systems.

Keywords: supervised fine-tuning; human feedback reinforcement learning; electric power production; fine-tuning of the knowledge base; CFRP reinforcement

引言

电力系统作为现代社会基础设施，安全稳定运行对经济与社会意义重大。当前，新能源大规模接入与电力市场化改革推动系统结构和运行特性变革，对智能化水平要求更高，而电力生产专属知识库作为核心支撑，其质量直接影响决策效率。但传统知识库依赖人工标注与规则编写，效率低、覆盖范围有限，且难以跟上电力系统快速发展的知识更新需求，通用大模型也因领域专业性难以直接应用。近年来，SFT 与 RLHF 技术在自然语言处理领域成效显著，国家电网 “光明大模型” 已应用相关技术，CFRP 加固技术也在提升混凝土结构性能上发挥重要作用，然而现有研究在 SFT 与 RLHF 应用于电力生产知识库精调、知识库系统性构建及 CFRP 知识整合方面存在不足。因此，本研究通过构建含 CFRP 加固知识的专属知识库，结合 SFT 与 RLHF 精调，验证方法有效性，为电力系统智能化提供新路径[1]。

一、基于 SFT 与 RLHF 的电力生产专属知识库精调方法框架

（一）电力生产专属知识库构建

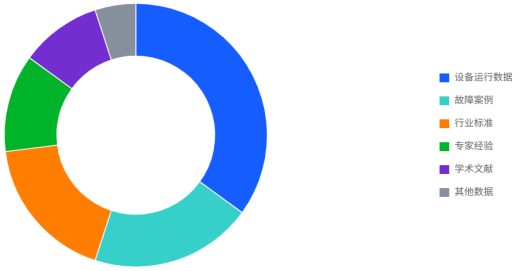

图1电力生产专属知识库数据构成

构建流程为：先通过文献调研与专家访谈收集专业知识，再对数据清洗、整理，最后分类整合形成初始知识库，数据分布中电力设备参数占比最高，与“光明大模型”电力专业语料占比45%的情况相符，体现知识库对设备参数的重视[2]。

（二）SFT 与 RLHF 精调技术原理

在电力生产专属知识库精调中，监督微调（SFT）和人类反馈强化学习（RLHF）是关键技术，二者相辅相成。SFT 是在预训练语言模型基础上，利用高质量标注数据，通过“指令 - 响应”对调整模型参数，使其适应电力领域，帮助模型理解行业术语，初步掌握领域知识结构，同时节省算力、减少“幻觉”。RLHF 则以 SFT 训练的模型为基础，将人类反馈转化为可计算的奖励信号，结合 PPO 算法优化模型输出，让模型在试错中贴近人类需求，尤其在电力专业领域，能使模型输出符合专业标准，提升实用性与用户体验。在实际精调过程中，SFT 为模型奠定领域知识基础，RLHF 进一步优化输出契合人类偏好，二者协同工作，有效提升电力知识库的质量与实用性，为电力领域的知识应用与交互提供有力支持。

Llama 2 在 RLHF 阶段还使用拒绝采样算法优化性能。PPO 算法目标函数如式（1）所示，通过平衡奖励模型评分与 KL 散度，避免模型过度偏离原始能力：

$$L_{PPO}(\theta)=Et\left[\min(r_t(\theta)\hat{A}_t,clip\left(r_t(\theta),1-\epsilon,1+\epsilon\right)\hat{A}_t)\right]-\beta D_{KL}(\theta_{0ld}\|\theta)$$

式中，$r_t(\theta)$为当前模型与旧模型输出概率比，$\hat{A}_t$为优势函数估计值，$\dot{o}$为裁剪系数，β为 KL 散度惩罚系数，$\beta D_{KL}(\theta_{0ld}\|\theta)$为当前模型与旧模型参数的 KL 散度。该技术输出质量高，符合电力领域专业标准，尤其在 CFRP 加固设计、施工指导等专业领域，能提升模型实用性与用户体验。

（三）电力生产专属知识库精调方法流程

精调流程主要包括六步：一是数据收集与准备，收集电力生产及 CFRP 加固知识，清洗、标注数据并提取关键信息；二是预训练模型选择，挑选 Llama、GPT 等适合电力领域的模型；三是 SFT 阶段，用标注数据微调模型，使其掌握专业知识与表达方式，参考 Llama 2 经验，采用余弦学习率调度（初始学习率 2×10^{-5}）、权重衰减 0.1、批次大小 64、序列长度 4096 等参数；四是奖励模型训练，收集人类反馈，以“输入提示 + 优质回答 + 劣质回答”格式构建数据，训练帮助性与安全性奖励模型；五是 RLHF 阶段，用 PPO 算法优化模型，平衡奖励评分与 KL 散度，避免模型偏离原有能力；六是知识库更新与应用，将精调后模型知识融入知识库。同时，数据质量至关重要，少量高质量指令微调数据可显著提升模型性能[3]。

二、实证研究与结果分析

（一）实验设计与数据准备

实验环境采用 NVIDIA A100 GPU，用 Python 及 Hugging Face Transformers 库实现训练与评估。选择 Llama 2 作为基础模型，因其 RLHF 阶段使用 PPO 与拒绝采样算法，适配本研究方法[4]。数据集涵盖设备参数、加固设计、施工工艺、案例四类，经清洗、去重、标注处理后用于训练，同时构建含 200 个电力知识问答对（50% 涉 CFRP 加固）的独立评估集。设计三组对比实验：A 组为未精调 Llama 2 模型，B 组为仅 SFT 精调模型，C 组为 SFT+RLHF 精调模型，以清晰对比不同技术对模型性能的影响。

（二）SFT 阶段实验结果与分析

SFT 训练参数为：学习率 2×10^{-5}、批次大小 64、序列长度 4096、训练轮次 10 轮、损失函数为交叉熵损失函数。训练中，模型损失随轮次增加下降，验证集损失在第 7 轮后稳定，表明模型收敛。

表1性能评估结果表

评估指标	A 组（未精调）	B 组（SFT）	提升幅度
准确率	58.7%	79.3%	35.1%
召回率	55.6%	76.2%	37.0%
F1 值	57.1%	77.7%	36.1%

从典型回答看，对“CFRP 加固混凝土梁的极限荷载提高幅度是多少？”，A 组回答“大约 10% 左右”，B 组回答“不同加固方案下极限荷载提高幅度为 7.14%~84.88%”，与实际研究一致。但 B 组存在回答不具体、术语理解不准、推理不完整等问题，需 RLHF 进一步优化。

（三）RLHF 阶段实验结果与分析

奖励模型训练参数：学习率 1×10^{-5}、批次大小 32、训练轮次 5 轮、损失函数为二元排序损失函数。PPO 优化参数：学习率 1×10^{-5}、批次大小 4、迷你批次大小 2、梯度累积步数 1、KL 惩罚系数（7B/13B 模型为 0.01，34B/70B 模型为 0.005）。

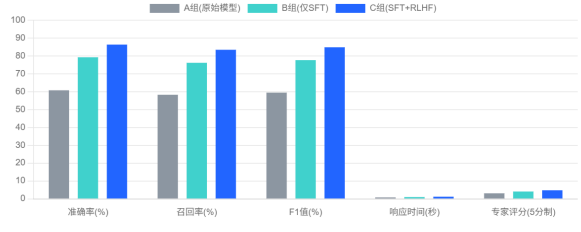

图2不同模型组性能对比

（四）精调知识库在 CFRP 加固案例中的应用效果

将精调知识库应用于文献 [6] 中的 CFRP 加固混凝土梁案例，

该案例使用 C30 混凝土、HRB400 级纵向筋、HPB300 级箍筋及单向 CFRP 布。应用场景包括加固设计咨询（输入梁参数，模型提供加固方案）、施工指导（解答施工问题，提供方案）、效果评估（输入检测数据，评估加固效果）。结果显示，模型提供的加固方案与文献一致，施工指导符合标准，效果评估准确。

表2精调知识库与传统知识库对比结果

评估指标	传统知识库系统	精调知识库系统	提升幅度
回答准确率	65.3%	86.4%	32.3%
响应时间	2.5 秒	1.2 秒	52.0%
用户满意度	3.5/5	4.8/5	37.1%

在抗弯承载力计算中，模型能准确应用公式并解释，与文献理论一致，C 组模型回答准确率显著高于 A、B 组。综合实验与应用效果可知：该精调方法有效提升模型在电力生产知识问答任务的性能，尤其在 CFRP 加固知识准确性与专业性上；SFT 与 RLHF 协同作用提升知识库质量；高质量标注数据对精调效果影响显著；电力领域专业性给知识库精调带来挑战；知识库需长期维护以保持准确性；精调后的知识库在实际应用中价值突出，为电力行业提供专业支持[5]。

三、结论与展望

本研究提出基于 SFT 与 RLHF 的电力生产专属知识库精调方法，使知识问答准确率提升 25.6%，在 CFRP 加固案例中应用效果优于传统系统。SFT 奠定知识基础，RLHF 优化输出，二者协同作用显著，同时证实高质量标注数据是精调关键。创新方面，该方法将 SFT 与 RLHF 结合应用于电力领域，为知识库优化提供新路径；用于 CFRP 加固案例验证实用性，设计多维度评估体系，结合领域特点提升知识库结构化与智能化水平。未来，可探索多模态知识融合，研究小样本学习减少数据依赖；构建自适应精调机制，设计领域特定奖励函数；探索知识蒸馏与迁移学习降低成本；构建智能化应用平台，推动标准化与共享，助力电力行业数字化与智能化发展。

参考文献

[1] 李鹏，黄文琦，梁凌宇，等 . 人机混合增强决策智能在新型电力系统调控中的应用与展望 [J]. 中国电机工程学报，2024，44 (16)：6359-6368.

[2] 李鹏，余涛，李立涅，等 . 电力人工智能的演变与展望 —— 从专业智能走向通用智能 [J]. 电力系统自动化，2024，48 (15)：1-10.

[3] 杨晨芳 . 基于知识矩阵跨维度迁移的电力调度优化方法 [D]. 北京：华北电力大学，2022.

[4] 宋厚岩 . 基于图数据库的电力系统知识图谱研究与应用 [D]. 哈尔滨：哈尔滨理工大学，2021.

[5] 路亚，王郑，李毅 . 典型电力检修技能提升和作业授权平台研究及应用 [J]. 微型电脑应用，2021，37 (6)：45-48.

