基于 LoRA 微调的大模型就业推荐框架

黄丽宏
广东省外语艺术职业学院，广东 广州 510310
DOI: 10.61369/TACS.2026080048

摘　　要： 校园招聘场景下，应届生岗位精准匹配是就业推荐的核心需求，传统小模型就业推荐算法存在精度不足、语义理解缺失、交互性差等缺陷。大语言模型（LLM）强大的自然语言理解能力，为解决该问题提供了新路径。本文提出一种融合协同过滤思想与低秩适应（LoRA）微调技术的大模型就业推荐框架，将学生结构化表格数据转为自然文本，通过协同过滤筛选潜在岗位，构建动态提示词（Prompt）并结合学生特征文本作为模型输入，依托 LoRA 微调实现就业岗位精准推荐。基于2021–2023年本校学生就业数据开展实验，结果表明，本文算法相较传统推荐算法及常规大模型微调推荐算法性能更优，准确率较传统算法最高提升0.2598，较通用大模型方法最高提升0.3930，可显著提升毕业生就业岗位推荐的精准度。

关 键 词： LoRA 微调；大语言模型；就业推荐；协同过滤；CFR-LoRA

Employment Recommendation Framework of Large Language Model Based on LoRA Fine-tuning

Huang Lihong
Guangdong Teachers College of Foreign Language and Arts, Guangzhou, Guangdong 510310

Abstract: In the campus recruitment scenario, the precise job matching for fresh graduates is the core demand of employment recommendation. Traditional employment recommendation algorithms based on small models have inherent defects such as low accuracy, insufficient semantic understanding and poor interactivity. With powerful natural language understanding capability, large language models (LLMs) provide a new solution to the above problems. This paper proposes a novel employment recommendation framework for large language models that integrates collaborative filtering and Low-Rank Adaptation (LoRA) fine-tuning technology. The framework converts students' structured tabular data into natural text, screens potential jobs through collaborative filtering, constructs dynamic prompts, and takes students' characteristic texts as model inputs to realize accurate job recommendation via LoRA fine-tuning. Based on the employment data of students in the college from 2021 to 2023, experimental results show that the proposed algorithm achieves better performance than traditional recommendation algorithms and conventional fine-tuned large model recommendation algorithms. Specifically, the accuracy is increased by up to 0.2598 compared with traditional algorithms and 0.3930 compared with general large model methods, which can significantly improve the accuracy of job recommendation for graduates.

Keywords: LoRA fine-tuning; large language model; employment recommendation; collaborative filtering; CFR-LoRA

引言

当前我国高等教育进入大众化阶段，高校毕业生规模持续扩大，大学生就业成为关乎社会经济发展与高等教育质量的核心民生问题。后疫情时代，高校毕业生“慢就业”现象日益凸显，该现象核心成因是毕业生无法匹配契合自身条件的理想岗位，多数应届生因求职匹配失败，最终选择暂缓就业。

现有就业推荐系统主要分为基于内容过滤、协同过滤、知识过滤及混合过滤四类方法，具备实现简单、可解释性强、低资源运行的优势，但存在明显短板：无法有效解决用户、岗位冷启动问题，对新应届生、新岗位难以生成有效推荐，且缺乏上下文语义理解能力，特征提取维度单一。大语言模型凭借优异的长文本处理、语义理解与特征挖掘能力，可精准捕捉学生求职的场景、偏好等隐性信息，弥补传统算法缺陷。

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

LoRA 微调即低秩自适应微调，依托模型低维内在维度假说，仅更新模型低秩参数矩阵，冻结大模型原始参数，既能适配下游专项任务，又可大幅降低显存与计算开销，实现高效微调。本文融合协同过滤与 LoRA 微调技术，结合动态 Prompt 工程构建就业推荐模型，依托相似学生岗位偏好一致性的核心逻辑，定制个性化动态提示词，针对性微调通用大模型，适配就业推荐专属场景。本文核心模型逻辑可概括为：$$y=x+PromptW+x+PromptBA$$，其中 $$y$$ 为最终推荐结果，$$x$$ 为学生特征文本描述，$$W$$ 为大模型原始参数，$$Prompt$$ 为动态提示词，$$A、B$$ 为 LoRA 低秩微调参数矩阵。

本文实验数据取自某高校2021-2023年学生就业数据集，为保护隐私，隐藏学生姓名并以编号替代。数据集包含岗位名称、职业类型、薪酬、院系专业、学历、专业岗位匹配度等核心字段。针对原始数据存在的缺失、表述不规范、冗余空格等问题开展预处理：采用列均值填充薪酬等少量缺失字段，删除大面积缺失的无效样本；统一专业、岗位的差异化表述；清除文本冗余空格，规范数据集格式。

一、模型设计

（一）模型总体框架

本文提出的 CFR-LoRA 就业推荐模型包含三大核心模块，分别为协同过滤模块、LoRA 微调模块与大模型推理模块，整体架构简洁高效、分工明确。其一，协同过滤模块基于学生特征相似度筛选相似用户，挖掘同类学生的就业岗位信息，为 Prompt 构建提供先验数据；其二，LoRA 微调模块通过引入低秩矩阵 A、B，对大模型注意力层与全连接层进行参数微调，实现通用大模型向就业推荐领域的适配；其三，大模型推理模块融合学生特征文本与动态 Prompt，依托微调后的模型参数，输出 Top-N 最优岗位推荐结果，同时剔除重复推荐岗位，保障推荐精准性。

（二）模型编码表示

传统推荐算法采用 multi-hot 编码处理学生结构化数据，存在编码稀疏、无语义关联、特征权重均等化、关键信息丢失等问题。本文摒弃传统编码方式，将学生结构化表格数据转化为自然语言文本描述，适配大模型输入特性，转换公式为 $$x_i=text(f_1,f_2,\dots,f_n)$$，式中 $$f_i$$ 为学生各项特征字段，$$text()$$ 为特征文本转换函数。

文本转换完成后，通过大模型分词器完成文本分词，得到 token 序列：$$tokens=Tokenizer(x_i)$$。对所有 token 进行稠密向量映射，叠加位置编码区分文本序列位置，依托大模型自注意力机制聚合全文上下文语义信息，最终将模型最后一层编码输出作为学生特征的全局向量表示，完整保留特征语义关联与权重差异。

（三）模型 Prompt 设计

基于协同过滤核心思想，特征相似的学生具备相近的岗位求职偏好。本文通过余弦相似度计算学生用户间的特征相似度：$$simU_{ij}=\cos(x_i,x_j)$$，其中 $$x_i、x_j$$ 为两名学生的特征向量表示。针对目标学生，筛选相似度最高的 Top-K 相似学生，提取对应学生的就业岗位信息，生成潜在推荐岗位序列 $$TopK_{irec}=[岗位1, 岗位2,\dots, 岗位 k]$$。

将 Top-K 相似学生岗位推荐先验信息与目标学生原始特征文本拼接，构建动态个性化 Prompt。该 Prompt 随学生个体特征与相似用户数据动态变化，具备极强的场景适应性，可为大模型推理提供精准的上下文约束，引导模型输出贴合学生需求的推荐结果。

（四）模型的微调设计

本文采用 LoRA 高效微调策略，冻结大模型原始参数矩阵 $$W_0$$，仅通过低秩矩阵 A、B 更新模型参数，参数更新公式为 $$W=W_0+\alpha BA$$，其中 $$\alpha$$ 为学习率，仅优化低秩参数增量 $$\Delta W=AB$$，大幅降低微调计算量。

模型以动态 Prompt 嵌入表示 $$h(prompt)$$ 为输入，通过更新后的模型参数完成推理，经 softmax 函数输出岗位预测概率分布，推理公式为 $$y=softmax(h(prompt)(W_0+BA))$$。微调过程采用交叉熵损失函数优化模型，损失公式为 $$L=-\sum y_i\log(y)$$，其中 $$y_i$$ 为学生真实就业岗位标签，通过迭代优化损失函数，持续提升模型推荐精度。

二、实验展示

（一）实验设置与基准模型

本文实验数据集取自某高职院校2021-2023年学生就业数据，数据集具备就业推荐场景典型的稀疏性特征，单名学生仅对应一个就业岗位。筛选清洗后得到4000名学生样本、1004个岗位样本，按照8:2比例划分训练集与测试集。实验采用中文适配性优异的 Qwen1.5 系列模型，设置多组参数规模（0.5B、1.8B、4B、7B）开展对比实验。

为验证本文 CFR-LoRA 模型的性能优势，设置多组传统算法与大模型微调算法作为基线模型：传统算法包括协同过滤推荐（CFR）、随机森林推荐（RFR）、多层感知机推荐（MLPR）、梯度提升推荐（GboostR）；大模型微调算法包括常规 LoRA、QLoRA、Adapter、BitFit。传统算法采用 multi-hot 特征编码，各类大模型算法统一采用本文的文本特征编码方式，保证实验变量唯一。

（二）评价指标

实验采用推荐系统通用评价指标，包括准确率（Accuracy）、精度（Precision）、召回率（Recall）、F1-Score，同时引入 Top-3场景下的 Precision@3、Recall@3及 NDCG 指标评估排序推荐性能，各指标计算公式均采用行业通用标准，全面衡量模型的预测准确性与排序合理性。

（三）实验结果与分析

传统算法对比实验结果表明，本文 CFR-LoRA 模型（Qwen1.5-1.8B）所有指标均最优。传统算法中 GboostR 性能最

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

优，但准确率仍低于本文模型9.08%，精度低5.34%；MLPR 准确率最低，CFR 整体综合性能最差，充分证明大模型语义挖掘能力相较于传统机器学习算法在就业推荐任务中具备显著优势，本文模型可有效捕捉学生与岗位的隐性匹配特征。同时，本文模型 Precision@3、NDCG 指标大幅领先传统算法，排序推荐性能更优异。

大模型算法对比实验显示，CFR-LoRA 在各参数规模 Qwen 模型上均优于 Adapter、BitFit、QLoRA 等主流微调算法。其中 Qwen1.5-1.8B 模型适配性最佳，性能达到峰值，模型参数规模与推荐性能并非正相关。7B 大模型因冗余特征维度过多，微调梯度噪声增加，导致收敛难度提升、性能出现波动。而本文模型在小参数模型下即可实现高精度推荐，具备高效、轻量化的优势，NDCG 指标在 Qwen1.5-4B 上达到0.8400，远超所有对比模型。此外，Adapter、BitFit、QLoRA 性能随模型规模提升呈正向波动，而本文模型稳定性更强、泛化性更好。

（四）消融实验结果

为验证各模块有效性，本文设置纯 Prompt、纯 LoRA、CFR-LoRA 三组消融实验。结果表明，单一模块均可实现基础推荐效果，且纯 LoRA 微调的性能显著优于纯 Prompt 推理，证明参数微调对大模型场景适配的核心作用。而融合动态 Prompt 与 LoRA 微调的 CFR-LoRA 模型性能全面超越单一模块，证明协同过滤构建的动态 Prompt 与 LoRA 微调存在正向协同效应。各模块相互适配、缺一不可，共同提升了模型的就业推荐精准度与稳定性。同时实验发现，纯 Prompt 性能不随模型规模增大而提升，进一步验证了专属微调与场景化 Prompt 结合的必要性。

三、结论

本文针对传统就业推荐算法语义缺失、精度低、适配性差等问题，提出一种融合协同过滤与 LoRA 微调的大模型就业推荐框架。通过将结构化学生数据文本化、基于用户相似度构建动态 Prompt、轻量化 LoRA 参数微调，实现了高校毕业生岗位的精准个性化推荐。多组对比实验表明，本文 CFR-LoRA 模型相较于传统机器学习算法与主流大模型微调算法，在准确率、召回率、F1-Score、NDCG 等核心指标上均存在显著优势。消融实验验证了动态 Prompt 设计与 LoRA 微调模块的协同增益作用。同时实验证明，就业推荐任务无需依赖超大参数模型，中小规模大模型经专项微调即可实现最优性能，为高校轻量化、高精度就业推荐系统的搭建提供了可行方案。未来可进一步优化相似度筛选策略，结合学生求职意愿动态权重，持续提升推荐的个性化适配效果。

参考文献

[1] 黄少成，唐雨晴 . 高校 " 慢就业 " 群体画像、成因及对策 [J]. 思想理论教育 ,2024(09):107-111.

[2] 张钦彤，王昱超，王鹤羲，等 . 大语言模型微调技术的研究综述 [J]. 计算机工程与应用 ,2024,60(17):17-33.

[3]Hamadi R. Large Language Models Meet Computer Vision: A Brief Survey[J]. arXiv:2311.16673v1,2023.

[4]Aghajanyan A,Gupta S,Zettlemoyer L. Intrinsic dimensionality explains the effectiveness of language model fine-tuning[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing.2021:7319-7328.

[5] 曾碧卿，陈鹏飞，姚勇涛 . 融合思维链和低秩自适应微调的方面情感三元组抽取 [J]. 计算机工程 ,2024,50(07):53-62.

[6]Mao Y,Ge Y,Fan Y,et al. A Survey on LoRA of Large Language Models[J]. arXiv:2407.11046v4,2024.

[7]Patel R,Vishwakarma S K. An Efficient Approach for Job Recommendation System Based on Collaborative Filtering[M].Advances in Intelligent Systems and Computing,2020,1077.

[8]Martinez-Gil J,Freudenthaler B,Natschläger T. Recommendation of job offers using random forests and support vector machines[C]//CEUR Workshop Proceedings,2018,2083:22-27.

[9]Sharma S,Rana V,Kumar V. Deep learning based semantic personalized recommendation system[J]. International Journal of Information Management Data Insights,2021,1(2):100028.

[10]Liu W,Fan H,Xia M. Credit Scoring Based on Tree-Enhanced Gradient Boosting Decision Trees[J]. Expert Syst. Appl,2022,189:116034.

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.