面向成人教育的科技信息智能分析与微调方法

王 晨[1] 戚玉涛[2]
WANG Chen QI Yutao

摘 要 针对成人教育数字化转型中存在的知识碎片化、资源更新滞后及个性化服务不足等问题，文章提出了一种基于低秩适应（LoRA）微调的科技信息资源智能分析方法，通过高效微调技术优化预训练模型，降低灾难性遗忘风险，实现科技信息资源与教育资源的深度整合。该方法有效支撑了跨学科知识关联检索与个性化资源推荐，为成人教育提供了可进化的智能支持系统，解决了成人教育中个性化学习需求、专业文献理解和资源更新等问题，推动成人教育服务范式从经验驱动向数据驱动和精准服务的方向转型。

关键词 科技信息资源；低秩适应；个性化推荐；知识图谱

doi：10.3969/j.issn.1672-9528.2025.11.034

0 引言

在数字化转型驱动下，成人教育正经历结构性变革，传统模式的时空限制、资源碎片化与个性化缺失问题亟待破解。当前在线教育平台虽突破时空边界，但其课程体系仍以离散知识点为核心，存在资源更新迟滞、技术嵌入不足等缺陷，难以支撑职业场景下的深度学习需求。科技信息资源作为科研创新的知识载体，其数字化文献在知网等平台呈现爆发式增长，但检索效率低和知识关联性弱成为其应用瓶颈。知识图谱技术为解决这一问题提供了新思路：通过构建结构化的知识网络，可实现科技信息资源的智能关联与动态更新，支持个性化知识推荐和精准检索，帮助学习者快速定位前沿研究成果，构建跨学科知识体系，显著提升知识迁移能力[1]。但在实际应用过程中，存在教育领域大模型适配问题，针对该问题，参数高效微调技术（parameter-efficient fine-tuning, PEFT）展现出独特优势[2]。相较于传统全参数调整的高成本与灾难性遗忘风险，PEFT 通过插入轻量化适配模块（Adapters）或采用低秩适应（LoRA）技术，在保留通用语义表征的同时，利用图结构信息实现领域知识的定向注入。通过构建教育专用术语本体库，结合图注意力机制与分层微调策略，形成课程知识库 - 科技信息资源库双向增强机制。通过语义推理实现教学资源的动态适配，构建可进化的智能教育生态，推动成人教育从经验驱动向数据驱动、从标准供给向精准服务的范式转型。

1 大语言模型微调技术

1.1 微调的必要性与挑战

随着知识更新速度的加快，传统的教学资源往往难以满足成人教育过程中对专业性和个性化的学习需求。通过构建课程知识库 - 科技信息资源库来更好地支持成人教育中的深度学习和跨学科的知识迁移。大规模预训练模型通过广泛的语料库训练，学习了丰富的语言特征和知识结构，但在面对如科技信息资源分析等专业化任务时，其通用性逐渐显现出局限性，导致无法满足特定领域的高精度需求，需要通过微调技术对预训练模型进行定制化优化[3]，提升其在领域特定任务中的适应性和准确性。

微调技术利用预训练模型已有的语言知识，并针对领域任务进行优化。例如，在计算机领域，微调后的模型能够有效识别与网络安全相关的术语和威胁模式，提高文献解析和任务执行效率。同样，在成人教育中，微调模型能够更精准地处理来自不同行业的专业文献，帮助学习者快速获取最新科技信息资源，提升学习效率。微调技术的优势在于能够显著提升模型在特定任务中的表现，特别是在机器学习、数据分析和跨学科知识应用等领域，经过微调的模型能够更精确地捕捉数据模式，提高预测和分类能力，而且能够在资源有限的情况下，快速实现领域适配，满足不同学习需求。

微调也存在资源消耗、灾难性遗忘和过拟合问题。当领域数据量不足时，模型可能丧失其通用性，影响多任务处理能力。微调过程中对超参数的敏感性也增加了实验复杂性，

1. 西安电子科技大学网络与继续教育学院 陕西西安 710071
2. 西安电子科技大学网络与信息安全学院 陕西西安 710071
[基金项目] 中国成人教育协会“十四五”成人继续教育科研规划 2023 年度课题“面向成人教育的科技信息资源分析与应用研究”（2023-888Y）

且微调效果严重依赖于数据集的质量与多样性，特别是在科技信息资源不断更新的环境中。为解决这些问题，引入低秩适应（LoRA）技术，通过对模型权重矩阵进行低秩分解，减少计算资源消耗，并通过冻结大部分原始模型参数，仅微调少量附加参数，降低灾难性遗忘的风险[4]。

1.2 低秩适应（LoRA）技术原理

针对大规模预训练模型在领域适应性和资源效率方面的问题，采用基于低秩适应（LoRA）方法的高效微调策略，以解决资源消耗、任务依赖性和灾难性遗忘问题。LoRA 通过将模型的关键参数矩阵分解为两个低秩矩阵，并仅更新这两个低秩矩阵的参数，从而实现高效微调，保持原始模型主干网络不变。这一方法能够在保留广泛领域知识的同时，针对特定任务进行定制化调整，特别适用于科技信息资源应用领域。

在多任务场景中，LoRA 表现出显著优势。例如，在自然语言处理与跨领域文本理解任务中，LoRA 能够有效调整模型对特定领域术语和句法结构的理解，确保模型在不同文本领域中保持一致的性能。通过低秩矩阵的引入，LoRA 大大降低了全参数更新的计算复杂度，适合资源受限环境下部署大规模预训练模型应用[5]。与传统微调方法相比，LoRA 能够在多任务系统中共享基础模型，通过独立调整低秩矩阵，避免重复训练主干模型，极大提高了训练效率。此外，LoRA 的参数冻结机制有效减少了灾难性遗忘的风险，确保了在模型适应新任务时不会丧失原有知识，特别适用于跨语言转换、特定领域机器翻译等应用场景。LoRA 通过高效的资源利用和灵活的任务适应，推动了大规模预训练模型在科技信息资源领域的高效应用和持续优化。以下是 LoRA 方法的关键公式描述：

（1）参数矩阵分解

假设模型中某一层的权重矩阵为：

$$\boldsymbol{W}_0 \in \mathbf{R}^{d\times k} \tag{1}$$

式中：d 是输入维度；k 是输出维度。

在 LoRA 方法中，权重矩阵 $\boldsymbol{W}_0$ 被分解为两个低秩矩阵，分别为：

$$\boldsymbol{A} \in \mathbf{R}^{d\times r} \quad \boldsymbol{B} \in \mathbf{R}^{r\times k} \tag{2}$$

式中：r 是分解的秩（通常$(r \ll \min(d,k))$），以满足关系：

$$[\boldsymbol{W}' = \boldsymbol{W}_0 + \Delta\boldsymbol{W} = \boldsymbol{W}_0 + \boldsymbol{A}\cdot\boldsymbol{B}] \tag{3}$$

式中：$\boldsymbol{W}'$ 是经过微调后的权重矩阵；$\Delta\boldsymbol{W} = \boldsymbol{A}\cdot\boldsymbol{B}$ 是在微调过程中更新的低秩矩阵。

（2）低秩矩阵更新

在微调过程中，仅更新低秩矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的参数，而主干网络的参数矩阵 $\boldsymbol{W}_0$ 保持不变。因此，微调的主要计算集中更新为：

$$[\Delta\boldsymbol{W} = \boldsymbol{A}\cdot\boldsymbol{B}] \tag{4}$$

有效减少了计算开销和内存使用。

（3）输入向量的变换

对于输入向量$\boldsymbol{x} \in \mathbf{R}^d$，通过 LoRA 微调后的输出向量$\boldsymbol{y} \in \mathbf{R}^k$可表示为：

$$[\boldsymbol{y} = \boldsymbol{W}'\cdot\boldsymbol{x} = (\boldsymbol{W}_0 + \boldsymbol{A}\cdot\boldsymbol{B})\cdot\boldsymbol{x}] \tag{5}$$

这表明微调后的输出不仅保留了原始模型的知识（通过 $\boldsymbol{W}_0$），还包括了新的适应性调整（通过 $\boldsymbol{A}\cdot\boldsymbol{B}$）。

（4）损失函数与优化

在微调过程中，损失函数 L 依赖于微调后的权重矩阵 $\boldsymbol{W}'$，并通过优化 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的参数来最小化：

$$L = L\left(\boldsymbol{W}'\cdot\boldsymbol{x}, \text{target}\right) \tag{6}$$

式中：L 是具体任务的损失函数。

（5）适应性与灾难性遗忘的控制

LoRA 通过保持 $\boldsymbol{W}_0$ 不变，减少了灾难性遗忘的风险，并且仅通过调整 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 来适应新任务：

$$[\boldsymbol{W}' = \boldsymbol{W}_0 + \boldsymbol{A}\cdot\boldsymbol{B}] \tag{7}$$

这种策略确保了模型能够在保持原有性能的同时，进行特定任务的定制化调整。

这些公式构成了 LoRA 方法的理论基础，展示了其在大规模预训练模型微调中的有效性。这种方法不仅减少了微调对原始模型的干扰，还通过简化参数更新策略，实现了在不影响整体结构的前提下逐步优化特定任务性能的效果，支持大规模预训练模型的领域扩展与应用。

2 基于知识图谱的大模型 LoRA 微调

在知识图谱相关的任务中，LoRA 的核心在于如何利用知识图谱中的结构化信息，通过低秩矩阵的引入进行领域微调，以增强模型在图谱推理和关系抽取中的表现。具体操作流程如下：

（1）知识图谱嵌入初始化

将知识图谱中的实体和关系转换为嵌入向量。通常采用图神经网络（GNN）或图注意力网络（GAT）来生成这些嵌入。在预训练阶段，这些嵌入通过与文本表示相结合，形成了基础的表示层。

（2）权重矩阵低秩分解

在模型的关键层（如自注意力层或前馈网络层），将权重矩阵进行低秩分解。假设原始权重矩阵为$\boldsymbol{W} \in \mathbf{R}^{d\times d}$，LoRA 将其分解为两个低秩矩阵$\boldsymbol{A} \in \mathbf{R}^{d\times r}$和$\boldsymbol{B} \in \mathbf{R}^{r\times d}$，其中 $r \ll d$。这种分解能够有效降低参数量，并专注于图谱相关信

息的微调。

（3）图谱关系矩阵引入

知识图谱中的关系矩阵（通常是稀疏的）被整合到低秩分解的权重调整中。具体实现中，通过将关系矩阵与 LoRA 的低秩矩阵相乘，确保图谱结构信息在模型的表示中得到保留。例如，实体关系的嵌入在矩阵 ***A*** 中进行投影，关系类型则在矩阵 ***B*** 中得到强化。

（4）分层调优

在 LoRA 的多层结构中，不同层次的低秩矩阵专门用于处理知识图谱的不同方面。例如，浅层的 LoRA 模块关注实体之间的直接关系，中间层侧重于复杂关系的推理，而深层则通过图结构信息优化最终的输出表示。这种层次化的设计有助于捕捉图谱中的全局信息和局部关系。

（5）微调与推理

在具体任务中，如关系抽取或知识补全，LoRA 仅更新低秩矩阵 ***A*** 和 ***B***，其余主干网络参数保持冻结。这种方式不仅降低了计算开销，还能确保在新任务微调时，保留原有的广泛领域知识，避免灾难性遗忘。推理时，模型使用调整后的权重进行知识推理任务，利用知识图谱中的结构信息进行更精准的关系预测。具体如表 1 所示。

通过知识图谱的微调，模型在处理科技信息领域问题时，能够更加深入地理解和推理科技概念、技术原理和模型结构，提供更准确、细致的答案。微调后的模型能准确抓住科学术语之间的联系，提升其在科技问答、关系推理中的表现。具体如表 2 所示。

表 1 科技信息知识图谱微调前后问答对比

问题	科技信息知识图谱微调前	科技信息知识图谱微调后
问题 1：神经网络和支持向量机的区别是什么？	神经网络和支持向量机都是机器学习算法。	神经网络是模拟人脑神经元连接的结构，适用于处理复杂的非线性问题，特别在深度学习中应用广泛。支持向量机（SVM）是一种线性分类模型，能够通过最大化分类边界实现高维数据的有效分类。
问题 2：Transformer 的主要创新点是什么？	Transformer 是一种深度学习模型。	Transformer 模型的主要创新点是引入了自注意力机制，它允许模型在没有序列处理限制的情况下捕捉句子中词语之间的依赖关系，大幅提升了自然语言处理任务的表现。
问题 3：深度学习在图像识别中的应用有哪些？	深度学习可以用于图像识别。	深度学习，特别是卷积神经网络（CNN），在图像识别中应用广泛，如用于自动驾驶中的目标检测、医学影像分析中的肿瘤检测等。它通过多层次特征提取，实现了从低级到高级的特征学习。
问题 4：如何通过知识图谱增强自然语言处理任务？	知识图谱可以提高文本理解能力。	知识图谱通过将实体和关系嵌入自然语言处理模型，使得模型能够利用结构化知识进行推理和生成，更好地理解上下文中的隐含关系，提升模型在命名实体识别、关系抽取和问答任务中的表现。

表 2 科技信息文档微调前后问答对比

问题	科技信息文档微调前	科技信息文档微调后
问题 1：大数据分析在科技信息领域的应用是什么？	大数据分析可以用于数据处理。	在科技信息领域，大数据分析可以用于处理海量的科研文献、专利数据和实验报告，帮助研究人员发现新的科研趋势、技术热点以及跨领域的创新机会。例如，通过文本挖掘可以自动生成科技领域的研究综述和引用分析。
问题 2：自动化科技文献摘要生成的主要挑战是什么？	自动生成摘要是一个难题。	自动化科技文献摘要生成的主要挑战包括准确提取关键信息、理解复杂的科技术语，以及生成连贯且符合领域规范的总结内容。此外，长文档的处理和多领域知识的集成也是生成摘要的难点。
问题 3：人工智能在科技文献分类中的优势是什么？	AI 可以进行文档分类。	人工智能通过深度学习模型，尤其是 Transformer 架构，能够自动识别科技文献中的主题、关键词和结构，精确地将文献归类到不同领域。相比传统分类方法，AI 在处理多维度信息和跨领域文献时表现更为出色。
问题 4：科技信息管理中的自动问答系统如何运作？	自动问答系统通过分析文本提供答案。	科技信息管理中的自动问答系统通过将用户提出的问题转化为向量，与科技文档的嵌入向量进行匹配，并通过大语言模型生成与问题相关的答案。该系统还利用知识图谱来增强问题解析与答案生成的准确性，尤其在处理复杂科技领域问题时表现优异。

3 基于 LoRA 的科技信息文档微调

在科技信息文档相关任务中，LoRA 的应用流程聚焦于如何针对领域文献进行定制化调整，以便模型在文档生成、领域术语识别及逻辑推理任务中表现出色。具体操作流程如下：

（1）领域文档嵌入与预处理

对领域文档进行预处理，提取出领域特有的术语、章节结构和引用关系。针对这些信息，生成专用的文档嵌入向量，以便在模型中与 LoRA 微调模块进行结合。这一步通常包括对文档的分段解析、引用关系图的构建以及关键术语的识别。

（2）权重矩阵低秩分解

在与知识图谱类似的方式下，对模型中用于文档处理的关键层进行低秩分解，生成低秩矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$。然而，针对文档的特化，LoRA 在分解过程中重点关注章节结构和逻辑层次的保留，使得微调后的模型能够更好地理解领域文档的语义组织。

（3）文档层次信息的融合

文档中的层次结构（如章节关系、段落逻辑）在 LoRA 的微调过程中以多级注意力的方式进行建模。具体来说，通过将段落位置编码与低秩矩阵相结合，确保模型能够在不同语义层次上捕捉文档中的逻辑递进关系。这种方法使模型在理解长文本或复杂技术文档时更加精准。

（4）基于领域术语的特化微调

LoRA 的低秩矩阵在特定领域术语的表示中进行优化，通过单独训练这些矩阵以调整模型对领域特有表达的敏感度。例如，在专利文献或学术论文中，LoRA 能够通过微调使模型更好地捕捉技术概念和专业术语的含义，提升生成和推理的准确性。

（5）文档生成与优化

在生成任务中，LoRA 通过其调整后的权重矩阵，在输出层中引导模型生成符合领域规范的文本。这一过程结合了领域术语的精确表达和文档逻辑的一致性，确保输出内容在形式和内容上均符合领域要求。此时，LoRA 通过控制低秩矩阵的更新，避免对模型主干网络的干扰，实现稳定且高效的文档生成。

LoRA 在知识图谱和科技文档的领域微调中，利用低秩矩阵分解和模块化调整的优势，分别在图谱推理和文档生成任务中实现了高效、资源友好的微调方案。这种方法在领域特化任务中展现了良好的适应性，既能确保模型在不同领域任务中的一致性，又能在资源受限环境中进行灵活部署。

微调后的模型在处理科技文献时，能够更好地生成准确、连贯的文档摘要，并处理复杂的文献结构和术语。同时，模型在大数据分析、文献分类和自动问答系统中的表现得到了显著提升。

4 结语

本文围绕成人教育背景下科技信息资源的智能分析与高效应用，针对大规模预训练模型在特定领域适应性与资源效率方面的突出挑战，提出了一种基于低秩适应（LoRA）的参数高效微调策略。LoRA 通过对模型关键权重矩阵的低秩分解，仅更新小规模附加参数，实现在计算资源受限条件下对模型的快速适配与灵活部署，同时显著降低了灾难性遗忘风险。结合知识图谱推理与科技文献处理任务，验证了 LoRA 在成人教育中构建智能化学习支持系统的实际价值。通过定制化优化模型对科技信息资源的理解与生成能力，使其更好地服务于成人学习者对专业性强、更新迅速的知识内容的个性化获取需求，提供更具适应性、智能化和可持续的成人教育科技信息资源应用途径 [6]。

参考文献：

[1] 于伟 , 王忠军 . 面向科技信息服务的人工智能技术应用 [J]. 中国科技信息 , 2021(10):68-70.

[2] 秦董洪 , 李政韬 , 白凤波 , 等 . 大语言模型参数高效微调技术综述 [J]. 计算机工程与应用 ,2025,61(16):38-63.

[3] 吴春志 , 赵玉龙 , 刘鑫 , 等 . 大语言模型微调方法研究综述 [J]. 中文信息学报 , 2025,39(2):1-26.

[4] 朱永康 , 高彦杰 . 基于 LoRA 及其变体的通用大语言模型微调方法 [J]. 上海电力大学学报 , 2025,41 (1):90-95.

[5]HAN S Y, WANG M Y, ZHANG J L, et al. A review of large language models: fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges[J]. Electronics, 2024, 13(24):5040.

[6]CHEN Y, CHEN H P, SU S Z. Fine-tuning large language models in education[C]//2023 13th International Conference on Information Technology in Medicine and Education (ITME), Piscataway: IEEE, 2023:718-723.

【作者简介】

王晨（1981—），男，陕西西安人，博士，高级工程师，研究方向：网络信息处理、大语言模型技术、现代远程教育技术等。

戚玉涛（1981—），男，河南潢川人，博士，教授，研究方向：人工智能安全、工业互联网安全、大数据与知识计算等。

（收稿日期：2025-06-13 修回日期：2025-11-04）