

电力物资供应链电子数据信任强度自适应分级方法

俞春华¹, 郭庆雷^{2,3}, 杨珂⁴, 柏能¹, 黄宜华⁵

1. 国网江苏省电力有限公司 物资分公司, 江苏 南京 210036;

2. 国网区块链科技(北京)有限公司, 北京 100077;

3. 国网区块链应用技术实验室, 北京 100077;

4. 国网数字科技控股有限公司, 北京 100077;

5. 计算机软件新技术国家重点实验室 南京大学 计算机学院, 江苏 南京 210023

摘要

电力物资供应链电子数据管理需求复杂, 而传统电子数据信任强度分级方法受限于静态规则与单一评价维度, 缺乏领域知识引导, 难以满足精细化与差异化的应用要求。针对这一问题, 提出一种电力物资供应链电子数据信任强度自适应分级方法(Adap-LLM), 将专家构建的多维评分体系与分级规则以结构化提示词形式注入大语言模型, 并结合参数高效微调实现领域知识深度融合, 实现对多种类型电力文档进行信任强度的自适应分级。实验结果显示, 该方法在真实电力及公共数据集上的各项指标均优于现有模型方法, 充分验证了本方法在电力物资大数据分级管理时的有效性与先进性, 具有良好的电力应用价值。

关键词

供应链电子数据分级; 大语言模型; 领域知识; 提示词工程; 参数高效微调

中图分类号: TP391

文献标志码: A

Adaptive grading method for trust intensity of electronic data in the electric power material supply chain

YU Chunhua¹, GUO Qinglei^{2,3}, YANG Ke⁴, BAI Neng¹, HUANG Yihua⁵

1. State Grid Jiangsu Electric Power Co., Ltd., Materials Branch, Nanjing Jiangsu 210036, China;

2. State Grid Blockchain Technology (Beijing) Co., Ltd., Beijing 100053, China;

3. State Grid Blockchain Application Technology Laboratory, Beijing 100053, China;

4. State Grid Digital Technology Holding Co., Ltd, Beijing 100077, China;

5. State Key Laboratory for Novel Software Technology at Nanjing University, School of Computer Science, Nanjing 210023, China

Abstract

The management of electronic data in power material supply chains is complex. Traditional methods for grading trust intensity are limited by static rules and single evaluation dimensions, and lack guidance from domain knowledge, making them insufficient for fine-grained and differentiated applications. To address this, we propose Adap-LLM, an adaptive grading method that injects expert-constructed multi-dimensional scoring systems and grading rules into large

language models through structured prompts, and further integrates domain knowledge via parameter-efficient fine-tuning. Experiments on real-world power datasets and public benchmarks show that the proposed method outperforms existing approaches across multiple metrics, demonstrating its effectiveness and potential application value in hierarchical management of power supply chain data.

Key words

Supply Chain Electronic Data Grading, Large Language Model (LLM), Domain Knowledge, Prompt Engineering, Parameter-Efficient Fine-Tuning

1 引言

随着能源互联网与电力供应链数字化的深度融合，电力物资供应链中的电子数据总量与类型不断增长与复杂化。对其可信性进行分级管理，直接关系到电网物资管理的审计合规性、合法性及运营效率。不同电力物资供应链电子数据在敏感性、法律效力和业务影响程度上存在差异。这种差异性需要在数据存储、查询与溯源的过程中，根据数据的信任强度等级实施自动化分级管理、差异化信息保护。

在电子数据自动分级与可信评估研究中，现有方法主要可分为规则驱动方法、深度语义建模方法和可信数据治理方法。规则驱动方法通常依赖人工规则、统计特征、概率推理或源可靠性估计实现数据分类与可信度判断，具有一定可解释性，但对规则设计和人工特征依赖较强，难以适应复杂业务文档的语义变化^[1-2]。深度学习方法通过多标签神经网络、分层图卷积和长文档建模机制增强文本语义表示能力，能够缓解传统方法对人工特征的依赖，但对法律风险、业务价值、审计追溯等领域规则的建模仍不充分^[3-5]。面向多源异构数据和工业场景的研究进一步引入跨模态融合、领域本体和动态信任管理机制，为复杂数据分类、可追溯治理和多维可信评估提供了参考^[6-10]。总体来看，现有方法在

电子数据分级中仍存在评价维度单一、规则体系静态、领域知识与语义模型融合不足等问题，难以满足电力物资供应链电子数据精细化、差异化和自适应分级需求。

针对上述问题，本文提出 Adap-LLM 方法。该方法以四维评分体系和 C/P/R 分级规则为知识基础，通过结构化提示词注入评分维度、分级阈值、推理步骤和输出约束，并结合 LoRA 微调学习文档语义特征与信任强度等级之间的映射关系，实现电子数据的自适应分级。

本文主要贡献如下：

- 1) 构建了面向电力物资供应链电子数据的四维评分体系和 C/P/R 分级标准，为电子数据差异化管理提供规则依据；
- 2) 提出了融合结构化提示词与 LoRA 微调的 Adap-LLM 方法，实现领域规则与大语言模型语义理解能力的协同建模；
- 3) 在 PPSCD 和 GBD-TB 数据集上开展对比实验、消融实验和权重敏感性分析，验证了所提方法的有效性和稳健性。

2 任务描述

在电力物资供应链电子数据信任强度自适应分级任务中，给定电力供应链中的电子文档集合为 $D = \{d_1, d_2, \dots, d_n\}$ ，其中每个文档 d_i 可表示为：

$$d_i = \{\text{content}_i, \text{metadata}_i, \text{context}_i\} \quad (1)$$

其中, $content_i$ 为文档主要文本内容, $metadata_i$ 包含文档类型、文档名、文档路径等元数据, $context_i$ 为文档所处业务环节、流程节点及上下文信息。任务目标是学习映射函数:

$$f: D \times K \rightarrow L_1 \times L_2 \quad (2)$$

其中, K 表示领域专家知识集合, 包括评分标准、分级阈值、典型案例等专家经验, $L_1 = \{C, P, R\}$ 表示一级分类标签, $L_2 = \{C_i, P_j, R_k | i, j, k \in \{1, 2, 3\}\}$ 表示二级分类标签。本文以 C/P/R 一级信任强度等级作为主要分级目标, 二级分类用于刻画一级等级内部的业务属性差异, 并为后续领域知识注入、模型推理约束和解释生成提供细化依据。

为保证分级规则具有明确的业务依据, 本文依据国网数据管理规范、物资供应链业务制度、合同履行管理要求及审计合规要求, 构建面向电力物资供应链电子数据的多维度量评分框架。该框架由国网物资领域专家组结合合同签订、招标采购、履约验收等典型业务环节归纳形成, 重点

刻画不同电子文档在法律效力、业务控制、审计追溯和敏感保护中的作用差异。在此基础上, 从法律约束、业务影响、审计追溯和敏感保护四个方面提炼出法律风险程度、业务重要性、审计合规要求和数据敏感性四个核心维度。

设 $s_j(d_i)$ 表示文档 d_i 在第 j 个评分维度上的得分, w_j 表示对应评分维度的权重, 则文档 d_i 的综合信任强度评分定义为:

$$S(d_i) = \sum_{j=1}^4 w_j \cdot s_j(d_i) \quad (3)$$

其中, 法律风险程度、业务重要性、审计合规要求和数据敏感性的权重分别设为 35%、30%、20% 和 15%。该权重设置综合考虑了法律效力、业务影响、审计追溯和敏感保护在电力物资供应链数据治理中的相对重要性, 并经领域专家讨论确定。

基于上述综合评分结果, 本文构建 C/P/R 三级信任强度分级标准, 如表 1 所示。在一级分类基础上, 本文进一步设置二级分类细则, 用于描述同一信任强度等级内部的业务属性差异, 具体如表 2 所示。

表1 电子数据信任强度一级分类标准与阈值

一级分类	评分阈值	分类特征	说明
C级	$S(d_i) \geq 75$	决定供应链成败的核心文档, 具有不可替代性和决定性影响	对供应链成败起关键作用
P级	$50 \leq S(d_i) < 75$	维护供应链流程运转的关键文档, 具有重要的过程控制价值	在流程运转中具有重要的控制价值
R级	$S(d_i) < 50$	记录和支撑业务操作的辅助文档, 主要起备查和追溯作用	辅助支撑和备查追溯

3 基于大模型的电力物资供应链电子数据信任强度自适应分级方法

Adap-LLM 方法以四维评分体系、C/P/R 分级标准及二级分类细则为领域知识基础, 通过结构化提示词将专家规则、评分依据、推理步骤和输出格式嵌入模型输

表2 电子数据信任强度二级分类细则

一级分类	二级分类	分类条件	典型文档示例
C级	C1级(法律强制类)	法律风险 ≥ 85 分	正式合同、中标通知书
	C2级(财务决算类)	业务重要性 ≥ 85 分	重大付款凭证、结算文件
	C3级(履约确认类)	审计合规 ≥ 85 分	验收报告、履约证明
P级	P1级(技术规范类)	技术复杂度高、专业性强	技术规范书、专家意见
	P2级(商务流程类)	商务流程关键节点	招标文件、评标报告
	P3级(管理控制类)	管理控制重要环节	质量报告、风险评估
R级	R1级(数据统计类)	数据统计类,格式标准化	货物清单、统计报表
	R2级(流程记录类)	流程记录类,时间序列特征	会议记录、操作日志
	R3级(申请通知类)	申请通知类,交互传达功能	申请表单、通知公告

入,并结合LoRA参数高效微调学习电力物资供应链场景下文档特征与信任强度等二级之间的映射关系。方法整体架构如图1所示。

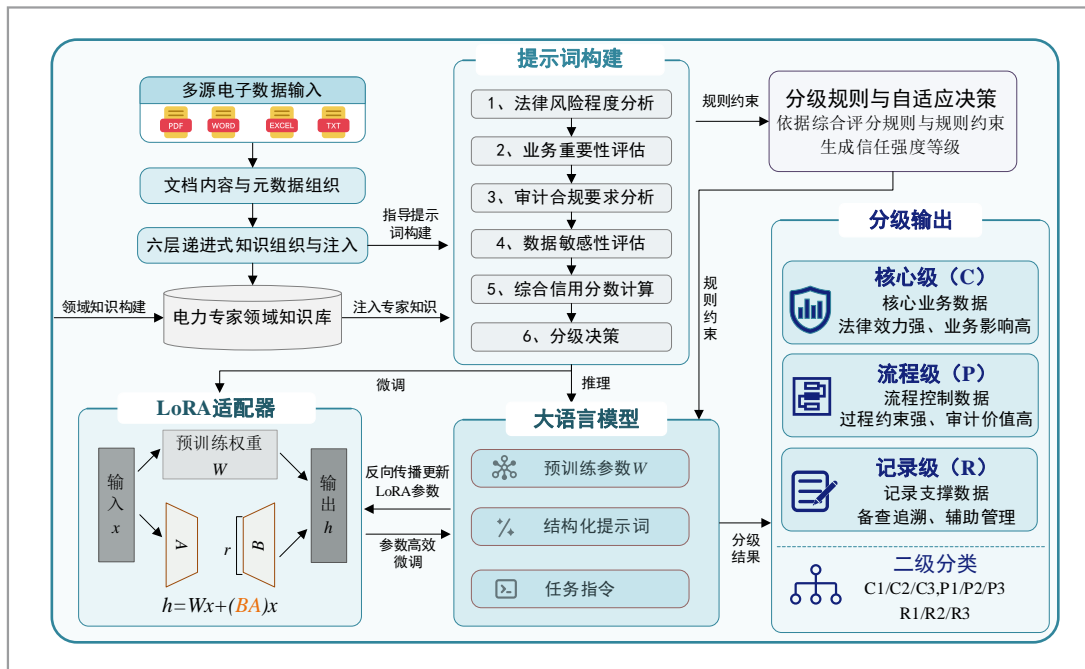


图1 方法整体架构示意图

3.1 领域知识注入的递进式架构设计

领域知识注入的目标是将电力物资供应链电子数据分级所涉及的业务规则、评

分标准和专家经验转化为可被大语言模型理解和遵循的推理约束,从而提升模型在专业场景下的判别一致性与结果可解释性。

针对电力物资供应链电子数据在法律效力、业务价值、审计追溯和敏感保护等方面的

差异，本文构建递进式领域知识注入架构，如图2所示。

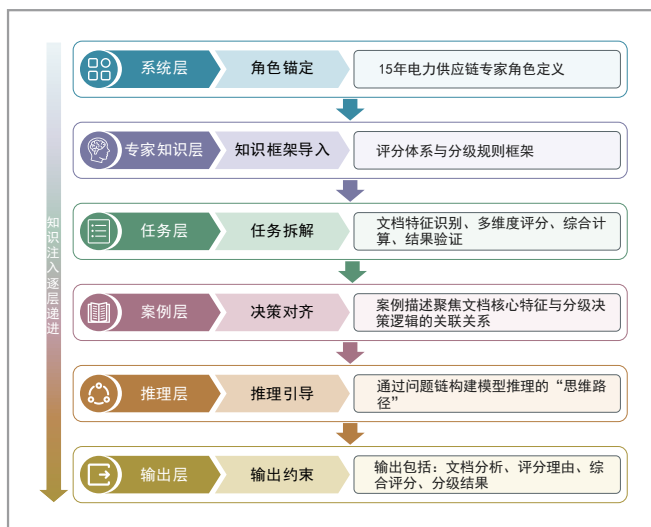


图2 递进式领域知识注入架构

该架构按照“角色限定—规则导入—任务分解—案例对齐—推理约束—格式规范”的逻辑组织领域知识，具体包括系统层、专家知识层、任务层、案例层、推理层和输出层。系统层用于限定模型的专业角色和任务场景；专家知识层引入四维评分体系、C/P/R分级标准及二级分类规则；任务层将分级过程分解为文档特征识别、维度评分、综合计算和等级判定；案例层通过典型样例建立文档特征、评分依据与分级结果之间的对应关系；推理层围绕文档类型、业务环节、法律效力、审计价值和敏感程度等要素约束分析路径；输出层统一模型生成结果的字段结构和表达形式。

通过上述设计，静态分级规则被组织为贯穿输入、推理和输出过程的知识约束链，使大语言模型能够在通用语义理解能力基础上融合电力物资供应链领域规则，为后续结构化提示词构建和参数高效微调提供一致的知识基础。

3.2 结构化提示词模板构建

在递进式知识注入架构基础上，本文进一步构建结构化提示词模板，将四维评分体系、C/P/R分级标准、二级分类细则、推理步骤和输出格式统一封装为模型输入。该模板由系统层提示词和用户层提示词两部分构成，如图3和图4所示。

系统层提示词主要承担领域规则注入功能，包括角色定义、核心任务、评分体系和输出约束。其中，角色定义用于设定模型的专业身份；核心任务明确模型需完成电子数据信任强度分级；评分体系嵌入法律风险程度、业务重要性、审计合规要求和数据敏感性四个维度及其权重；输出约束规定模型需按照统一结构生成文档特征、评分依据、综合得分和分级结果。

用户层提示词主要承担任务实例化功能，包括推理引导、待分析内容和案例指示。推理引导用于约束模型按照“文档类

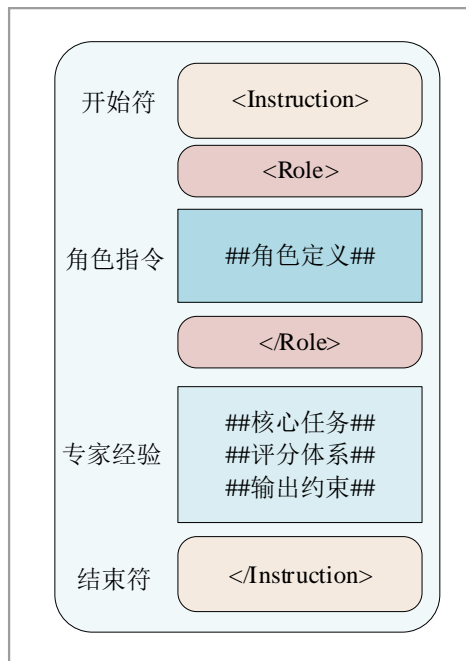


图3 系统层提示词构建框架

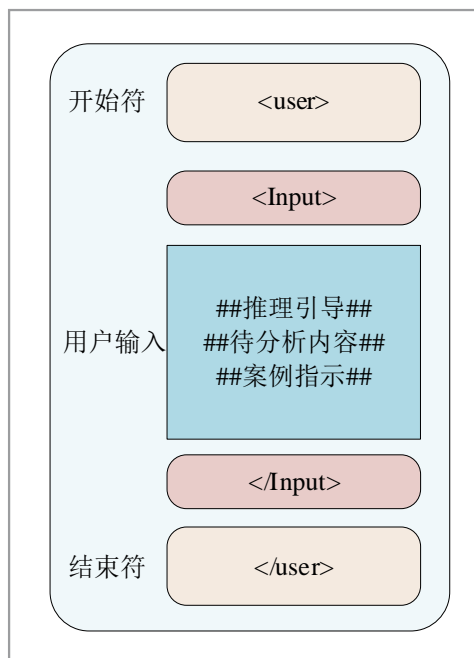


图4 用户层提示词构建框架

该设计通过显式化中间分析步骤增强模型对复杂分级任务的推理能力^[11]；待分析内容提供文档文本、元数据及业务上下文；案例指示通过典型样例增强模型对分级边界和业务语义的理解。

表3详细定义了领域知识注入的提示词模板具体构成。在模板的System部分，为Instruction的内容，明确声明了角色信息、专家知识、任务约束等；input部分提供具体的分析步骤要求、待评估文档以及案例指示。同时，该模板还规定了大语言模型的类型以及其相对应输入输出的界定符。

表3 模板参数示例表

参数名	参数值
System	<code>< im_start >instruction\n</code>
System	<code>< im_end >\n</code>
user_start	<code>< im_start >input\n</code>
user_end	<code>< im_end >\n</code>
llm_start	<code>< im_start >assistant\n</code>
llm_end	<code>< im_end >\n</code>

经过结构化提示词模板封装后，完整输入组织形式如图5所示。模型输入由`<Instruction>`和`<Text>`两部分组成，前者用于承载电力物资供应链电子数据的业务属性、评分维度、分级规则和输出约束，后者用于提供待分析文档内容及其元数据信息；模型输出则为包含文档特征、评分依据、综合得分和分级结果的结构化回复。通过上述组织方式，提示词不仅向模型提供分类任务指令，还将专家规则、推理路径和输出格式统一纳入输入上下文，使模型能够在领域规则约束下完成信任强度分级。

型识别—关键特征提取—四维评分—综合得分计算—等级判定”的顺序完成分析，

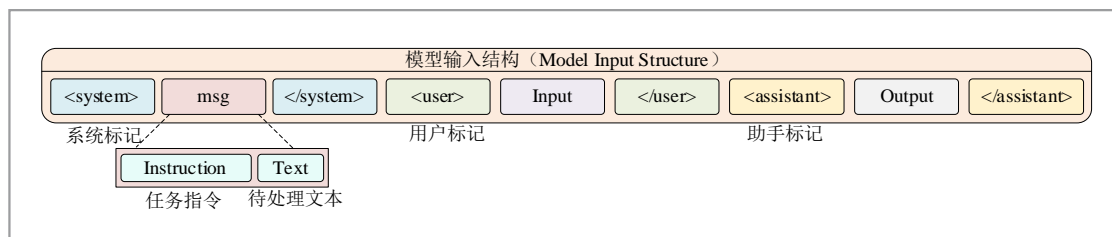


图5 面向指令微调的模型输入组织结构

3.3 面向分级任务的模型微调方法

为进一步提升模型对电力物资供应链场景的适配能力，本文在结构化提示词基础上采用 LoRA^[12] 参数高效微调方法。该方法冻结基座大语言模型主体参数，仅在注意力层和前馈网络中引入低秩适配矩阵，通过少量可训练参数学习电力物资文档特征、领域规则与信任强度标签之间的映射关系，从而在降低训练成本的同时保持模型原有语义理解能力。

设基座模型中待适配权重为 W_0 ，LoRA 通过低秩矩阵 A 和 B 对权重增量进行近似，则前向计算可表示为：

$$h = W_0 x + B A x \quad (4)$$

其中 x 为输入， W_0 在训练过程中保持冻结， A 和 B 为可训练低秩矩阵。该机制使模型能够在不进行全参数更新的条件下，学习电力物资供应链电子数据分级任务中的领域差异性特征。

本文采用指令监督微调范式构建训练样本。输入序列由结构化提示词、待分析文档及领域分级规则组成，目标输出包括文档关键特征、四维评分、评分依据、综合得分、一级分类和二级分类等字段。训练目标是在给定输入条件下最大化标准输出序列的生成概率，其损失函数定义为：

$$L = - \sum_{t=1}^T \log P_{\theta}(y_t | y_{< t}, X) \quad (5)$$

其中， θ 表示 LoRA 引入的可训练适配参

数， x 表示模型输入， y_t 表示目标输出序列中第 t 个 token。通过该训练过程，模型不仅学习文档内容与 C/P/R 等级之间的对应关系，还能够进一步内化提示词中嵌入的评分规则、推理步骤和输出格式约束。最终，本文方法可形式化表示为：

$$\text{Label} = \text{LLM}_{\text{LoRA}}(\text{Prompt}_{\text{Template}}, \text{Doc}) \quad (6)$$

其中， $\text{Pr ompt}_{\text{Template}}$ 表示结构化领域知识提示词， Doc 表示待分级电子文档， LLM_{LoRA} 表示经 LoRA 微调后的大语言模型， Label 表示模型生成的信任强度分级结果。

4 实验

4.1 数据集

本文实验使用私有电力物资供应链数据集 (Private Power Supply Chain Dataset, PPSCD) 和跨领域通用评估数据集 (General Business Document Trust Benchmark, GBD-TB) 进行验证。PPSCD 来源于电力物资供应链真实业务场景，覆盖合同签订、招投标、履约验收等关键业务环节。原始数据共包含 3856 份异构电子文档，格式包括 PDF、Word、Excel 和 TXT 等，典型文档类型包括正式合同、技术规范书、招标文件、验收报告、会议记录、统计报表等。为保证数据质量，本文对原始文档进行统一文本抽取、噪声

清洗、隐私脱敏和关键内容提取，形成用于模型训练和评估的核心文本内容。

在数据标注阶段，本文采用“双盲标注—一致性核查—争议复核”的流程生成数据集标签。具体而言，由未参与评分体系构建的国网物资领域专家依据四维评分体系和C/P/R分级规则对全部样本进行双人独立标注，并对标签不一致、四维评分差异超过设定阈值或综合得分接近C/P/R分界点的样本引入额外专家复核裁定，最终以复核结果作为标准标签。标注规则构建、样本标注与训练提示词设计相互分离，从而降低评分规则设计与样本标注之间的同源偏置风险。考虑到电力物资供应链电子文档普遍篇幅较长，单份文档中通常包含多个能够反映其信任强度的业务片段，本文在保持专家标签不变的前提下，对原始文档进行片段级样本构建，将标题、关键条款金额等内容组织为模型输入，并与对应专家标签组合形成“指令—输入—输出”结构化微调样本。最终，3856份原始电子文档被构建为18650条结构化样本。

为验证本文方法的跨领域泛化能力，本文进一步基于公开数据源构建跨领域评估数据集GBD-TB。该数据集融合ChnFinAnn、CAIL和THUCNews三个公开中文数据源，分别覆盖金融公告、司法文书和新闻文本。本文依据C/P/R分级规则对其进行标签映射：将具有强法律效力或高风险约束属性的司法文书、股权冻结及股权质押公告映射为C级；将反映企业关键业务流程和监管披露要求的股权回购、股权增持及股权减持公告映射为P级；将主要承担信息传递、事实记录和备查参考功能的财经简讯、事务性通知及一般新闻文本映射为R级。

PPSCD与GBD-TB数据分布如表4所示。为分析模型在临近分级阈值样本上的判别能力，本文进一步对PPSCD数据集中的边界样本进行统计。根据表1所示C/P/R分级阈值，将综合评分位于[58,62]区间的样本定义为R/P边界样本，将综合评分位于[73,77]区间的样本定义为P/C边界样本，统计结果如表5所示。

表4 PPSCD与GBD-TB数据分布

数据集	分类	文档类型	数量
PPSCD	C	合同、中标通知书、结算文件、验收报告等	6133
PPSCD	P	技术规范书、招标文件、评标报告、质量报告等	7322
PPSCD	R	会议记录、统计报表、货物清单、通知公告等	5195
合计	—	—	18650
GBD-TB	C	判决类文书、股权冻结公告、股权质押公告	4000
GBD-TB	P	股权回购、股权增持、股权减持公告	4000
GBD-TB	R	财经简讯、事务性通知、一般新闻文本	4000
合计	—	—	12000

4.2 评测标准

本文使用准确率（Acc）、精确率

（Pre）、召回率（Recall）和F1值作为电力物资供应链电子数据信任强度自适应分

表5 PPSCD数据集边界样本分布统计

数据集	样本类型	C类	P类	R类	数量	占PPSCD比例/%
PPSCD	R/P边界样本	0	810	740	1550	8.31
PPSCD	P/C边界样本	780	910	0	1690	9.06
PPSCD	全部边界样本	780	1720	740	3240	17.37
PPSCD	非边界样本	5353	5602	4455	15410	82.63
合计	—	6133	7322	5195	18650	100.00

级任务的评估指标。准确率表示预测类别与真实类别一致的样本数占总样本数的比例。精确率、召回率和F1值采用 one-vs-rest 方式，分别对各类别进行计算，即在计算某一类别指标时，将该类别视为正类，其余类别视为负类。其中，精确率表示被预测为该类别的样本中实际属于该类别的比例，召回率表示实际属于该类别的样本中被正确预测为该类别的比例，F1值为精确率与召回率的调和平均值。本文进一步对各类别的精确率、召回率和F1值进行宏平均，作为最终的 Pre、Recall 和 F1 指标。各类别指标计算所涉及的 TP、FP、TN 和 FN 定义如表6所示。

表6 分类任务评估指标定义表

	预测为正类	预测为负类
实际正类	TP	FN
实际负类	FP	TN

则每个指标的公式如下：

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Pre} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (10)$$

4.3 实验设置

本文以 Qwen3-4B 作为主要基座模型，并引入 LLaMA、DeepSeek 和 ChatGLM3 等大模型进行对比实验。实验环境为 Linux 操作系统，两台 NVIDIA T4-16G GPU 服务器，编程语言为 Python 3.10.8，深度学习框架为 PyTorch 2.6。

为避免同一原始文档的不同片段同时出现在训练集和测试集中，本文先在原始文档层面按照 8 : 1 : 1 划分训练集、验证集和测试集，再分别对各子集内文档进行片段级样本构建，并采用分层抽样保持类别比例基本一致。模型微调采用 LoRA 方法，冻结基座模型参数，仅更新适配参数。LoRA 作用模块包括 q_proj、k_proj、v_proj、o_proj、gate_proj、up_proj 和 down_proj；秩参数 r 设为 16，缩放系数 α 设为 32，Dropout 设为 0.05。优化器采用 AdamW，学习率设置为 $5e-5$ ，最大文本长度设置为 2048，受显存限制，训练批次大小为 1，推理温度系数设为 0.1，以降低生成随机性并提高输出稳定性。

本文以 C/P/R 一级信任强度等级作为主要评价对象，其中 C 级、P 级和 R 级分别对应核心级、流程级和记录级电子数据，用于表征电力物资供应链电子数据的主体信任强度等级；而二级分类为各主体等级

下的细化子类，用于刻画同一等级内部的业务属性差异，并作为领域知识注入、推理约束和结果解释生成的细粒度规则依据。实验过程中，通过规则匹配抽取模型输出中的C/P/R一级分类字段，并与人工标注

的主体信任强度等级进行比对，最终计算各项评价指标。实验所用提示词示例如表7所示，该模板将专家定义的四维评分体系、分级阈值以及推理逻辑显式地注入到模型的输入上下文文中。

表7 训练提示词示例

<pre> "instruction": 【角色定义】 你是一位具有电力供应链管理经验的专家,熟悉供应链风险控制与合规管理。 【核心任务】 根据专家经验和标准化评分体系,对供应链电子文档进行信任强度分级。 【专家经验评分体系】 文档信任强度评估采用四维评分法: 1)法律风险程度(权重35%):…… 2)业务重要性(权重30%):…… 3)审计合规要求(权重20%):…… 4)数据敏感性(权重15%):…… 【综合评分公式】 总分=法律风险程度×0.35+……+数据敏感性×0.15。 【分级标准】 C级:总分≥75分…二级分类包括C1、C2、C3、P1、P2… "input": 【分析步骤】 1)识别文档类型和关键信息; 2)按四个维度进行专家评分; 3)计算综合得分并确定分级; 4)提供分析依据。 【待分析文档】 XXXXX "output": { XXXX } </pre>

4.4 对比实验

为验证本文方法的有效性与跨模型适配能力，本文在GBD-TB和PPSCD两个数据集上选取传统模型、预训练模型、纯

大模型方法和本文方法进行对比。其中，传统与预训练基线包括Longformer^[13]、LSTM^[14]、BiLSTM-Attention^[15]和MacBERT^[16]，分别用于表征序列建模、注意力增强序列建模、中文预训练语义建

模和长文本建模能力。纯大模型方法采用 Prompt-only 设置, 在不进行参数微调的情况下, 调用 LLaMA^[17]、DeepSeek-LLM^[18]、ChatGLM3^[19]和 Qwen3^[20]系列中可获得的高性能 API 版本, 并输入完整结构化提示词完成推理。本文方法则选取与

上述模型系列对应的可本地部署轻量基座模型, 包括 LLaMA-7B、DeepSeek-LLM-7B、ChatGLM3-6B、Qwen3-4B, 在其基础上引入领域知识注入与 LoRA 微调, 构建 Adap-LLM 系列模型。各模型实验结果如表 8 所示。

表 8 本文方法和对比模型的指标值

模型	GBD-TB(公共数据集)				PPSCD(私有数据集)			
	Acc	Pre	Recall	F1	Acc	Pre	Recall	F1
传统模型								
LSTM	79.54	78.12	77.65	77.88	86.74	85.92	84.53	86.15
BiLSTM-Attention	83.41	82.76	81.94	82.35	89.06	88.42	87.95	88.18
MacBERT	86.34	85.71	85.16	85.43	90.72	90.18	89.64	89.91
Longformer	87.21	86.45	85.92	86.18	91.25	90.47	89.81	90.83
大模型方法(Prompt-only)								
LLaMA	87.86	87.24	86.71	86.91	91.82	91.26	90.74	91.00
DeepSeek-LLM	89.42	88.95	88.36	88.65	93.18	92.64	92.11	92.37
ChatGLM3	88.73	88.14	87.62	87.88	92.56	92.07	91.48	91.77
Qwen3	90.16	89.68	89.13	89.40	93.74	93.21	92.76	92.98
本文方法(Adap-LLM+不同大模型基座)								
Adap-LLM _{LLaMA}	91.18	90.62	90.21	90.41	94.63	94.08	93.66	93.87
Adap-LLM _{DeepSeek}	92.03	91.46	91.18	91.32	95.46	94.92	94.37	94.64
Adap-LLM _{ChatGLM3}	91.56	91.02	90.85	90.93	95.12	94.58	94.25	94.41
Adap-LLM _{Qwen3}	92.45	91.88	91.67	91.77	95.83	95.20	94.61	95.48

由表 8 可知, 传统序列模型整体性能相对有限, 引入双向建模与注意力机制后, BiLSTM-Attention 相较于 LSTM 取得一定提升, 说明注意力机制有助于捕捉文档中的关键片段。相比之下, MacBERT 和 Longformer 进一步提升了分级性能, 表明预训练语义表示和长文本建模能力对电力物资供应链电子文档分级具有积极作用。大模型 Prompt-only 方法整体优于传统模型和预训练模型, 说明大语言模型在强提示条件下已具备较好的语义理解和规则推理能力。但由于 Prompt-only 未经过领域任务微调, 其对电力物资供应链业务语境

和边界样本的适配能力仍然有限。本文方法在不同基座模型上均较对应 Prompt-only 基线取得提升, Qwen3-4B 取得最佳结果, 说明在规则约束下的中文业务文档分级任务中, 模型性能不仅取决于参数规模, 还与中文语义理解能力、指令遵循能力、结构化输出稳定性以及领域提示模板适配程度密切相关。上述结果表明, 结构化领域知识注入与 LoRA 微调能够进一步增强模型对分级规则、业务语义和输出格式的适应能力, 且该方法不依赖单一基座模型, 具有较好的跨模型迁移性。此外, 各模型在 PPSCD 上的表现普遍高于 GBD-

TB, 主要由于PPSCD与本文任务场景和标签规则更加一致, 而GBD-TB包含金融公告、司法文书和新闻文本等跨领域数据, 语义分布更复杂, 分级难度更高。

大模型训练与微调的效率不仅取决于模型参数量, 还受到算力系统、并行加速、存储系统及容错机制等基础设施因素影响。大模型基础设施建设应兼顾性能、可靠性和可扩展性, 训练过程中的检查点访问性能也会影响系统恢复效率与资源利用率^[21-22]。因此, 本文在模型性能之外补充统计Token消耗、端到端推理时间、输出吞吐量和峰值显存, 以从系统开销角度评估方法可部署性。其中, API模型的峰值显存不可观测; 本地Adap-LLM模型部署于NVIDIA T4环境, 显存占用由推理过程记录获得。

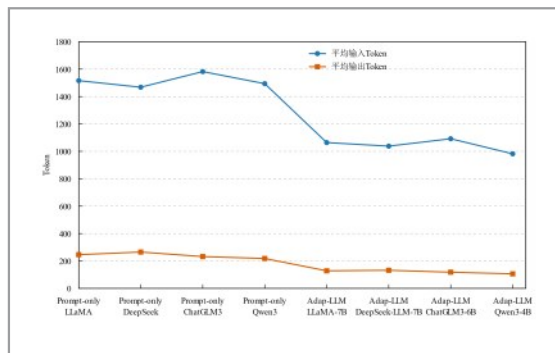


图6 Token消耗

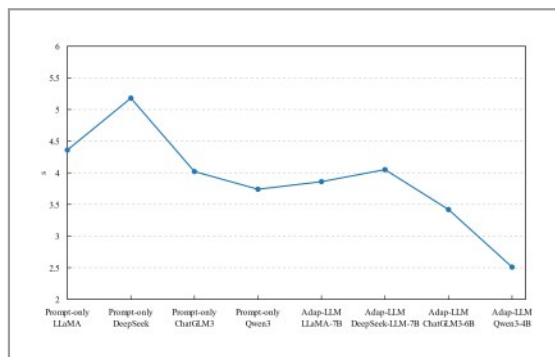


图7 平均推理时间

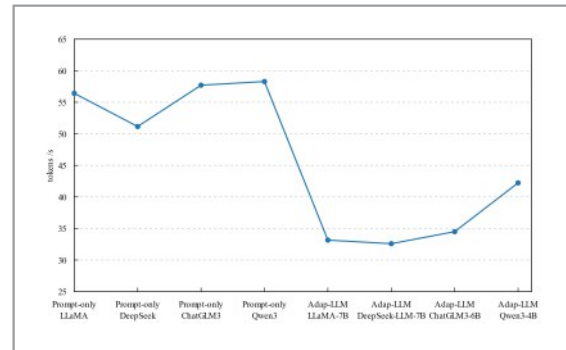


图8 输出吞吐量

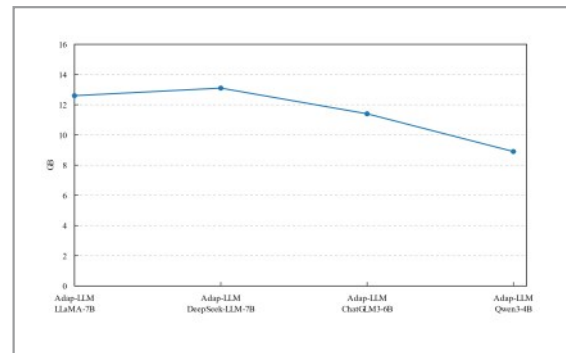


图9 峰值显存

4.5 消融实验

为验证领域知识注入、分步推理引导与LoRA微调对分级性能的影响, 本文以Qwen3-4B为基础模型, 在GBD-TB与PPSCD两个数据集上开展消融实验。实验设置如下:

1) Base: 未微调Qwen3-4B, 仅输入基本任务说明, 不引入领域知识和结构化约束;

2) Knowledge Prompt: 在Base基础上仅加入四维评分体系、C/P/R分级规则和结构化输出约束, 但不引入分步推理引导、典型案例示范和LoRA微调, 用于验证基础领域规则提示对模型分级性能的影响;

3) LoRA w/o Knowledge: 采用不含

四维评分体系和C/P/R显式规则的训练样本进行LoRA微调，但保留统一标签输出格式，仅学习文档内容与分级标签之间的映射关系；

4) Rule Prompt w/o Reasoning: 在

LoRA微调中加入四维评分体系和分级规则，但去除分步推理引导；

5) Adap-LLM: 采用本文完整方法，同时引入领域知识提示词、分步推理引导、结构化输出约束和LoRA微调。

表9 消融实验分析结果

模型	领域知识	分步推理	结构化输出	LoRA	GBD-TB(公共数据集)				PPSCD(私有数据集)			
					Acc	Pre	Recall	F1	Acc	Pre	Recall	F1
Base(基线)	×	×	×	×	83.12	82.45	81.56	82.35	88.42	87.68	86.95	87.91
Knowledge Prompt	✓	×	✓	×	86.27	85.64	84.93	85.28	90.73	90.12	89.56	89.83
LoRA w/o Knowledge	×	×	✓	✓	89.54	88.92	88.15	88.65	92.74	91.89	91.25	92.12
Rule Prompt -w/o Reasoning	✓	×	✓	✓	91.36	90.81	90.24	90.52	94.52	93.86	93.41	93.63
Adap-LLM	✓	✓	✓	✓	92.45	91.88	91.67	91.77	95.83	95.20	94.61	95.48

由表9可知，Base模型在未引入领域知识和参数微调的情况下表现相对有限，表明通用大语言模型虽然具备一定的语义理解能力，但仅依赖预训练知识难以充分适配电力物资供应链电子数据的专业分级需求。与Base相比，Knowledge Prompt取得一定提升，说明将四维评分体系、C/P/R分级阈值和结构化输出格式显式注入提示词后，能够增强模型对分级依据和业务语义的理解能力。然而，该设置未进行参数微调，模型在面对边界样本和跨领域样本时，仍存在规则遵循不稳定的问题。LoRA w/o Knowledge明显优于Base和Knowledge Prompt，说明参数高效微调能够有效学习文档语义特征与分级标签之间的映射关系，提升模型对电子文档分级任务的适配能力。但由于缺少显式领域知识引导，模型主要依赖训练样本中的统计相关性进行判别，对法律风险、业务重要性、审计合规要求和数据敏感性等分级依据的表达仍不充分。Rule Prompt w/o

Reasoning在LoRA微调基础上进一步引入四维评分体系和分级规则后继续提升，表明领域知识与参数微调具有较好的互补作用。领域规则能够为模型提供更明确的判别边界和业务约束，但由于该设置去除了分步推理引导，模型在复杂文档和临界分数样本上仍可能出现理由不充分或分类依据不一致的问题。本文完整方法Adap-LLM在两个数据集上均取得最优结果，与LoRA w/o Knowledge相比Adap-LLM分别提升3.12和3.36个百分点；与Rule Prompt w/o Reasoning相比，分别提升1.25和1.85个百分点。实验结果表明，领域知识提示词、结构化输出约束、分步推理引导和LoRA微调能够共同提升模型的分级性能。其中，领域知识提示词增强了模型对分级规则和业务语义的理解，LoRA微调提升了模型对领域数据分布的适配能力，分步推理引导则进一步提高了模型在复杂样本和边界样本上的判别稳定性，验证了本文所提Adap-LLM方法设计

的有效性。

4.6 四维权重敏感性分析与边界样本性能分析

为检验四维权重配置的稳健性，本文在PPSCD验证集上开展单因素敏感性分析。以法律风险程度、业务重要性、审计合规要求和数据敏感性的基准权重(0.35, 0.30, 0.20, 0.15)为参照，每次仅扰动一个维度权重，其余维度按原始比例归一化，保持权重和为1。实验中固定验证样本、

四维单项评分、模型推理方式及C/P/R分级阈值，扰动范围设为0.10~0.50，步长为0.05，并以Macro-F1评价分级性能。

由图6可知，各维度权重曲线在基准权重附近均保持较高水平，说明本文权重配置具有较好的稳定性。其中，法律风险程度和业务重要性对结果影响相对更明显，审计合规要求发挥稳定的辅助判别作用；数据敏感性在权重过高时性能下降，表明其更适合作为补充维度参与综合评分。上述结果验证了四维权重设置的合理性。

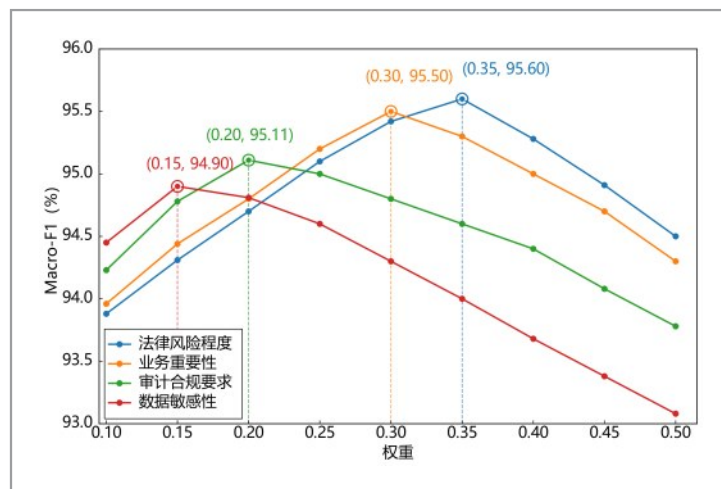


图10 四维评分权重敏感性分析结果

为进一步考察模型在临近分级阈值样本上的判别能力，本文针对表5中统计得到的PPSCD边界样本开展性能分析。与权重敏感性分析不同，边界样本分析不改变四维权重和C/P/R分级阈值，仅在基准权重配置下统计不同样本集合的Acc与

Macro-F1。由图7可知，边界样本性能低于非边界样本，说明临界样本判别难度更高，但本文方法在边界样本上仍保持90%以上的Acc和Macro-F1，表明所提方法对分级阈值附近样本仍具有较好的判别能力。

任强度分级需求，提出了一种融合领域知识与大语言模型的自适应分级方法Adap-LLM。该方法基于电力物资供应链业务规则构建四维评分体系和C/P/R层次化分级标准，并通过结构化提示词与LoRA微调

5 结论

本文面向电力物资供应链电子数据信

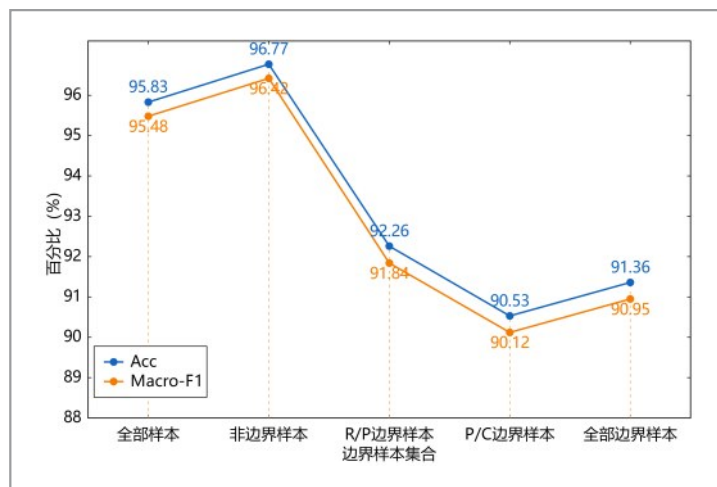


图11 PPSCD数据集边界样本性能分析

实现领域知识注入，使模型能够结合文档语义特征和专家规则完成信任强度判别。实验结果表明，本文方法在真实电力物资供应链数据集和跨领域公共数据集上均取得较优性能，能够有效提升电子数据分级的准确性和稳定性。进一步地，跨模型对比实验验证了所提方法对不同大语言模型基座的适配能力，消融实验验证了领域知识注入、分步推理引导和LoRA微调等模块的有效性，权重敏感性分析验证了四维评分权重配置的稳健性。研究结果说明，将专家分级规则、结构化推理约束与参数高效微调相结合，能够为电力物资供应链电子数据的自动化、精细化和可解释分级提供有效支撑。后续将进一步引入证据片段定位、规则触发链和检索增强生成机制^[23]，以提升分级结果的可审计性和跨场景泛化能力。

参考文献：

[1] HUANG ZHI, LIU PENG, ZHANG ZHIZHEN, et al. Automated Classification and Grading of Power System Data -

base Fields Based on Rule Association and Probabilistic Reasoning[C]//2024 10th IEEE International Conference on High Performance and Smart Computing (HPSC). NYC, NY, USA: IEEE, 2024: 83-88.

[2] Li Q, Li Y, Gao J, et al. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation[C]//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. Snowbird Utah USA: ACM, 2014: 1187-1198.

[3] SUN Y, QI L, XU G. Classification of Extra-High Voltage Project Documents Based on Improved Multi-Label Neural Network[C]//2023 5th International Conference on Applied Machine Learning (ICAML). Dalian, China: IEEE, 2023: 202-208.

[4] LIU T, HU Y, GAO J, et al. Hierarchical Multi-Granularity Interaction Graph Convolutional Network for Long Document Classification[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024, 32: 1762-1775.

- [5] LIU T, HU Y, WANG B, et al. Hierarchical Graph Convolutional Networks for Structured Long Document Classification [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34 (10): 8071–8085.
- [6] DAI S, SONG D, WANG B, et al. CCEndNet: Cross-Modal Cascaded Encoder-Decoder Network for Multisource Data Fusion Classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1–23.
- [7] K R, SRINIVAS H, S S. Industrial information extraction through multi-phase classification using ontology for unstructured documents [J]. Computers in Industry, 2018, 100: 137–147.
- [8] ZHANG Y, MA Z, LUO S, et al. Dynamic Trust-Based Redactable Blockchain Supporting Update and Traceability [J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 821–834.
- [9] YANG T, ZHANG Y, KANG C, et al. Exploration of Data Classification and Grading Strategy for Power Grid Enterprises [C]//Proceeding of the 2024 5th International Conference on Computer Science and Management Technology. Xiamen Fujian China: ACM, 2024: 1332–1336.
- [10] NITTI M, GIRAU R, ATZORI L. Trustworthiness Management in the Social Internet of Things [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1253–1266.
- [11] WER J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2022.
- [12] HU E J, SHEN Y, WALLIS P, et al. LoRA: Low-Rank Adaptation of Large Language Models [EB/OL]. (2021–10–16) [2025–09–30]. <http://arxiv.org/abs/2106.09685>.
- [13] BELTAGY I, PETERS M E, COHAN A. Longformer: The Long-Document Transformer [EB/OL]. (2020–12–02) [2025–09–30]. <http://arxiv.org/abs/2004.05150>.
- [14] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735–1780.
- [15] WANG Z, YANG B. Attention-based Bidirectional Long Short-Term Memory Networks for Relation Classification Using Knowledge Distillation from BERT [C]//2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCOM/CyberSciTech). Calgary, AB, Canada: IEEE, 2020: 562–568.
- [16] CUI Y, CHE W, LIU T, et al. Pre-Training With Whole Word Masking for Chinese BERT [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 3504–3514.
- [17] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and Efficient Foundation Language Models [EB/OL]. (2023–02–27) [2026–04–28]. <http://arxiv.org/abs/2302.13971>.
- [18] DEEPSEEK-AI, BI X, CHEN D, et al. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism [EB/OL]. (2024–01–05) [2026–04–28]. <http://arxiv.org/abs/2401.02954>.
- [19] ZENG AH, XU B, WANG BW, et al. ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tools [EB/OL]. (2024–07–30) [2026–04–

- 28]. <http://arxiv.org/abs/2406.12793>.
- [20] YANG A, LI A, YANG B, et al. Qwen3 Technical Report[EB/OL]. (2025-05-14) [2025-09-30]. <http://arxiv.org/abs/2505.09388>.
- [21] 郑纬民. 构建支持大模型训练的计算机系统需要考虑的4个问题[J]. 大数据, 2024(1): 1-8.
- Weimin ZHENG. Four issues to consider in building a computer system supporting large model training[J]. Big data research, 2024, 10(1): 1-8.
- [22] 滕云, 张广艳, 孙大为, 等. 深度学习模型训练过程检查点访问性能优化方法[J]. 大数据, 2026(2): 75-84.
- Teng Yun, Zhang Guangyan, Sun Dawei, et al. Checkpoint accessing performance optimization method for the deep learning model training process[J]. BIG DATA RESEARCH, 2026, 12(02): 75-84.
- [23] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2020.

作者简介



俞春华 (1973-), 男, 本科, 高级经济师, 主要研究方向为供应链管理、数字化转型。

收稿日期: XXXX-XX-XX

通信作者:

基金项目: 国家电网有限公司科技项目资助 (5700-202418240A-1-1-ZN)

Foundation Items: