

基于教育数据分类分级的大模型框架研究

孙志磊¹, 张旗², 赵澄¹

1. 浙江工业大学 信息技术中心, 浙江 杭州 310023

2. 杭州安恒信息技术股份有限公司, 浙江 杭州 310051

DOI: 10.61369/TACS.2026050055

摘 要 : 随着教育数字化转型的深入推进, 智慧校园与教育大数据系统产生的海量多源异构数据, 对数据治理的安全性、合规性与智能化水平提出严峻挑战。传统人工规则或关键词匹配方法存在泛化能力弱、语义理解不足等局限, 难以适配教育场景的动态演化特征。为此, 本文提出面向教育领域的可解释智能数据分类分级框架 EDCF-LLM (Educational Data Classification and Grading Framework with LLM)。该框架将预定义分类体系建模为有向树结构, 引入智能体 (Agent) 作为流程控制器, 通过分层提示策略引导大语言模型 (LLM) 构建从粗粒度到细粒度的递归推理路径; 设计多轮决策校验机制抑制语义漂移, 并融合检索增强生成 (RAG) 机制缓解 LLM “幻觉” 风险。实验结果表明, 该方法在准确率、召回率与 F1 分数上显著优于传统基线模型, 实现高精度自动分类与推理路径可追溯, 为教育数据安全管控与可信共享奠定技术基础。

关 键 词 : 教育数据治理; 分类分级; 大语言模型; 递归推理; 检索增强生成

Research on Large Model Framework Based on Education Data Classification and Grading

Sun Zhilei¹, Zhang Qi², Zhao Cheng¹

1. Information Technology Center, Zhejiang University of Technology, Hangzhou, Zhejiang 310023

2. Hangzhou DBAPP Security Co., Ltd., Hangzhou, Zhejiang 310051

Abstract : With the in-depth advancement of educational digital transformation, the massive multi-source heterogeneous data generated by smart campuses and educational big data systems has posed severe challenges to the security, compliance, and intelligence level of data governance. Traditional manual rule-based or keyword matching methods have limitations such as weak generalization ability and insufficient semantic understanding, making it difficult to adapt to the dynamically evolving characteristics of educational scenarios. To address this, this paper proposes an interpretable intelligent data classification and grading framework for the educational field — EDCF-LLM (Educational Data Classification and Grading Framework with LLM). The framework models the predefined classification system as a directed tree structure, introduces an Agent as the process controller, and guides the Large Language Model (LLM) to construct a recursive reasoning path from coarse-grained to fine-grained through a hierarchical prompting strategy. A multi-round decision verification mechanism is designed to suppress semantic drift, and the Retrieval-Augmented Generation (RAG) mechanism is integrated to mitigate the "hallucination" risk of LLMs. Experimental results show that this method significantly outperforms traditional baseline models in precision, recall, and F1-score, achieving high-precision automatic classification and traceable reasoning paths. It lays a technical foundation for the security control and trusted sharing of educational data.

Keywords : educational data governance; classification and grading; Large Language Model (LLM); recursive reasoning; Retrieval-Augmented Generation (RAG)

引言

在教育信息化纵深发展背景下, 学生行为轨迹、学业表现、心理健康记录等多元数据快速积累, 成为精准教育决策的核心生产要素。此类数据蕴含敏感个人信息与机构核心资产, 泄露将引发隐私侵犯与社会风险, 亟需建立科学的数据治理体系。数据分类与分级作为核心环节, 分别承担资源组织与安全防护职能: 分类依据语义特征构建结构化目录, 分级根据敏感度设定差异化保护策略。

当前教育领域数据分类分级实践仍依赖人工规则或关键词匹配技术,面临三大局限:一是难以覆盖非结构化文本隐含语义;二是规则库更新滞后于政策与业务演化;三是缺乏深层语义建模能力,误判率高。部分研究引入人工神经网络、TF-IDF 权重等特征工程手段^[1,2],但性能依赖特征设计质量;基于 BERT 的预训练模型虽在自然语言处理任务中表现突出^[3],却因教育领域术语密集、标注数据稀缺,面临领域适配不足与过拟合问题。

为此,本文提出 EDCF-LLM 框架,将 LLM 的语言理解优势与结构化知识体系的逻辑约束相融合,通过智能体驱动的分层递归推理机制,实现灵活且可控的自动化分类分级,解决现有方法在语义理解深度、动态适应能力与决策透明度方面的不足。

一、相关工作

(一) 国内外数据分类分级实践

数据分类分级范式已从“合规驱动”演进至“智能治理”。国际上,欧盟 GDPR 确立“设计即隐私”原则,实现个人数据全链条管控;我国《数据安全法》实施后,形成国家—行业—组织三级标准体系。教育领域中,刘英群等^[4]提出“数据资产生态”概念,甘清云等^[5]论述分类分级的安全基础作用,张聪^[6]构建高校数据分类模型,杨伟平等^[7]提出二维分类分级框架。但现有研究多聚焦静态体系设计,缺乏动态语义演化建模,传统“规则—匹配”范式在非结构化数据、多模态场景下泛化能力弱,亟需实现从“人工定义”向“语义理解驱动”的跃迁。

(二) LLM 相关应用

LLM 凭借卓越的上下文理解与少样本泛化能力,在自然语言处理领域展现变革潜力。Wei 等^[8]提出 FLAN 模型,验证指令微调的零样本泛化能力;Farr 等^[9]提出多模型共识不确定性量化方法,提升 LLM 辅助标注可信度;Kuzman 等^[10]构建 LLM 教师—学生框架,实现无人标注的多语言分类;Hong 等^[11]融合机器学习与 LLM,提升异构数据威胁检测性能。

然而,LLM 直接用于数据治理面临三重挑战:一是“幻觉”问题威胁结果准确性;二是知识固化难以适配外部规范更新;三是决策过程黑箱化,缺乏可解释性,无法满足教育数据治理的合规要求。因此,需引入结构性引导与外部知识干预,构建可信智能治理系统。

二、EDCF-LLM 框架设计

EDCF-LLM 框架的核心是构建智能体 (Agent)、RAG 与 LLM 三位一体的协同体系,形成“控制—执行—知识支撑”闭环。框架将 LLM 转化为特定语境下的合规推理决策支持引擎,整体结构如图 1 所示。

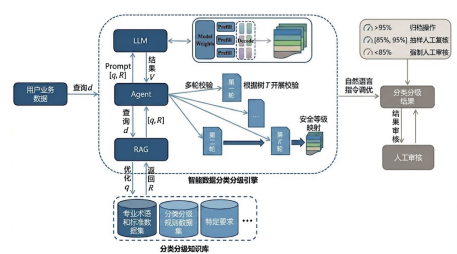


图 1 EDCF-LLM 模型框架

(一) Agent 流程控制中枢

Agent 贯穿 LLM 的知识建构与推理执行阶段,通过显式建模分类结构、编排提示序列、约束推理空间及验证决策路径,弥补 LLM 在记忆保持、逻辑一致性与过程可控性上的缺陷。

1. 学习构建阶段: 分层认知路径形式化

Agent 将预定义分类分级框架 F 解析为有向树结构 $T=(V,E)$, V 为分类标签, E 为父子隶属关系,根节点 r 代表顶层维度,叶节点 L 映射至敏感度等级 $S=\{L1,L2,L3\}$ 。基于该结构,Agent 生成多轮提示序列:

$$Prompt_t = [SysPrompt; Context_t; Query], t=1,2,\dots,k \quad t=1,2,\dots,k \quad (1)$$

其中 k 为最大推理深度,引导 LLM 渐进式认知:首轮识别根节点子节点完成粗粒度判定,中间轮聚焦当前节点子类细化,最终轮定位叶节点并输出敏感等级。

2. 推理执行阶段: 路径规划与动态验证

面对待分类数据 d ,Agent 规划从根节点到叶节点的推理路径,每步决策需满足 $(v_{i-1}, v_i) \in E$,防止语义漂移或非法跳转。若 LLM 返回无效节点,模型自动回退并调整提示重试。

3. 结果输出与置信度评估

成功构建分类路径 $Path(d) = (v_1, v_2, \dots, v_k)$ 后,通过映射函数 ϕ 确定敏感等级 $S(d) = \phi(v_k)$,输出结构化元组:

$$Output(d) = (d, Path(d), s, Conf(d)) \quad (2)$$

其中置信度 $Conf(d)$ 定义为各步最大概率的均值:

$$Conf(d) = \frac{1}{k} \sum_{k=1}^k \max_{c \in children(v_{i-1})} P(c | d, v_1, \dots, v_{i-1}, T) \quad (3)$$

用于人工审核排序与模型诊断。

(二) LLM 语义理解引擎

LLM 作为核心语义理解引擎,在 Agent 控制下输出候选类别置信度分布,不独立决策。通过“控制—执行”分离架构,Agent 维护推理状态、构建上下文并限定候选集合,LLM 专注语义解码,协同完成分类路径构建。

1. 多步递归分类推理过程

给定数据实例 d ,目标是在有向树 T 约束下识别最优路径 $\pi = (v_1, v_2, \dots, v_k)$,整个过程分解为 k 轮递归推理,第 i 步决策为条件概率最大化问题:

$$v_i = \arg \max_{c \in children(v_{i-1})} P(c | d, v_1, \dots, v_{i-1}, T, K_i) \quad (4)$$

具体分为三阶段:首轮粗粒度范畴判定,聚焦顶层语义匹配;次轮子类细化,收缩语义空间增强上下文敏感性;最终轮末端分类与敏感度判定,形成完整语义路径。

2. 检索增强生成 RAG

为缓解 LLM 领域适配不足问题，RAG 作为动态知识供给引擎嵌入推理流程。推理前，RAG 依据当前上下文从知识库 K（含结构化规范与非结构化示例）检索相关片段，注入 Prompt 引导 LLM 精准判别。

工作流程如下：首先对原始查询 q 重生成 q' ，提升检索效力；随后采用混合检索策略融合关键词检索（BM25 算法）与向量检索（余弦相似度）：

$$BM25(q', d) = \sum_{w \in q'} IDF(w) * \frac{f(w, d)(k_1 + 1)}{f(w, d) + k_1 \left(1 - b + b * \frac{|d|}{avgdl} \right)} \quad (5)$$

$$Sim(q', d) = \frac{Embed(q')^T * Embed(d)}{\|Embed(q')\| * \|Embed(d)\|} \quad (6)$$

$$Score_{q', d} = \alpha * BM25(q', d) + (1 - \alpha) * Sim(q', d), \alpha \in [0, 1] \quad (7)$$

检索所得 Top-n 片段与 q' 拼接输入 Agent，确保 LLM 推理基于可验证知识，抑制幻觉。

三、实验设计与性能评估

（一）实验数据

实验数据来源于智慧校园平台日志与业务系统记录，涵盖教师、学生、教学、科研等完整业务场景。数据经脱敏处理后，通过规范化清洗（格式标准化、关键字段对齐、语义归一化），依据《教育数据分类分级指南》细化为 18 个子任务集合，含 26350 个有效字段样本。

（二）对比实验

选取三类基线方法对比：规则匹配器（Rule-based Matcher）、SVM+TF-IDF 模型、微调 BERT 模型（Fine-tuned BERT），分别代表人工规则、传统机器学习与深度语义建模技

术。从时间效率、准确率、召回率、F1 分数四维度评估性能。

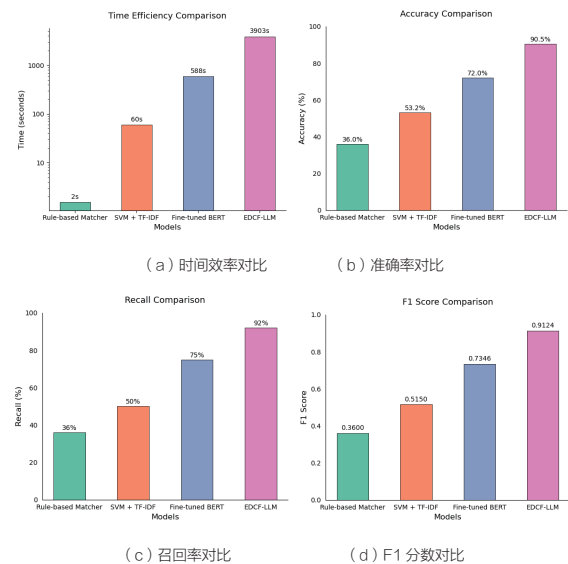


图 2 各模型综合性能与效率对比

如图 2 所示，EDCF-LLM 在准确率、召回率与 F1 分数上显著超越基线模型：相比规则匹配与机器学习模型，性能提升尤为明显；即使与微调 BERT 相比，仍保持更高的准确率与召回率，体现出上下文建模与复杂模式捕捉的优越性。

四、结论

本文提出 EDCF-LLM 框架，融合 LLM 语义理解能力与结构化知识体系约束机制，通过有向树结构建模分类逻辑、Agent 驱动分层递归推理、RAG 增强知识适配，实现教育数据智能化分类分级。实验结果表明，该框架在真实教育场景中性能优于传统方法，为教育数据安全治理、合规共享与价值释放提供有效技术路径。未来可进一步优化检索策略与推理效率，拓展多模态教育数据的适配能力。

参考文献：

- [1]Haisheng Gu, Qing Li, Duanming Shen. Research on sensitive text classification detection and classification based on improved artificial neural network[J]. Journal of Combinatorial Optimization, 2023(46): DOI: 10.1007/s10878-023-01085-8.
- [2]Sheridan Paul, Onsjö Mikael. The hypergeometric test performs comparably to TF-IDF on standard text analysis tasks[J] Multimedia Tools and Applications, 2023(83):28875-28890.
- [3]Hossain Md.Mithun, Hossain Md. Shakil , Mridha M. F. et.al. Multi task opinion enhanced hybrid BERT model for mental health analysis[J]. Scientific Reports, 2025: DOI: 10.1038/s41598-025-86124-6
- [4]刘英群, 周潜, 韩锡斌. 教育场景驱动的高校数据治理 [J]. 中国电化教育, 2024, 444: 38-43.
- [5]甘清云. 浅谈数据安全之数据分类分级 [J]. 网络安全技术与应用, 2025, 1: 67-69.
- [6]张聪. 高校数据分类分级策略的探讨与实践 [J]. 网络安全与数据治理, 2024, 43(6): 53-57.
- [7]杨伟平, 曾德华, 段婷婷, 赵显. 数字化转型背景下的教育数据分类分级研究 [J]. 现代教育技术, 2025, 1: 89-97.
- [8]Wei Jason, Bosma Maarten, Zhao Vincent Y. et al. Finetuned Language Models Are Zero-Shot Learners[J]. Computation and Language. 2021: arxiv-2109.01652.
- [9]David T. Farr, Lynnette Hui Xian Ng, Iain J. Cruickshank. et al. Ensemble-Based Uncertainty Quantification for Reliable LLM Classification in Social Data Applications [J]. IEEE Access. 2025: DOI: 10.1109/access.2025.3585414.
- [10]Taja Kuzman, Nikola Ljubešić. LLM Teacher-Student Framework for Text Classification With No Manually Annotated Data: A Case Study in IPTC News Topic Classification[J] IEEE Access, 2025 :DOI: 10.1109/access.2025.3544814
- [11]Chaerim Hong, Taeyeon Oh. Optimization for threat classification of various data types-based on ML model and LLM [J]. Scientific Reports. 2025: DOI: 10.1038/s41598-025-05182-y.