

基于多智能体专家协同的学生精准画像与学业规划系统构建与应用

Building a Multi-Agent Expert System for Personalized Student Profiling and Learning Plan

张瑶^{1*}, 蒋可², 屠炳豪³

¹²³ 浙江工业大学 教育学院

*202105720240@zjut.edu.cn

【摘要】针对当前教育评价中画像维度单一及单体大模型推理能力不足的问题，本研究基于学习者模型构建涵盖知识、认知、情感、行为的四维评估框架，并提出一种拟人化专家协同的多智能体教育架构。通过驱动不同角色智能体协作，系统实现从精准诊断到适性规划的闭环。实证分析表明，该系统能有效识别学生群体，并通过提示词驱动的动态优先机制整合建议形成个性化学业规划。评估结果显示，其生成方案在准确性、个性化与教育合理性上表现较优，为生成式 AI 赋能精准育人提供了新的视角。

【关键词】教育多智能体；学生画像；个性化学习；学习者模型

Abstract: To address the problems of single-dimensional student profiling and insufficient reasoning ability of monolithic large language models in current educational evaluation, this study constructs a four-dimensional evaluation framework covering knowledge, cognition, emotion and behavior based on the learner model, and proposes a multi-agent education architecture with expert collaboration. By driving collaboration between agents with different roles, the system realizes a closed loop from accurate diagnosis to adaptive planning. Empirical analysis shows that the system can effectively identify student groups and formulate personalized academic plans through prompt-driven dynamic priority integration mechanism. The evaluation results show that its generation scheme is excellent in accuracy, personalization and educational rationality, providing a new perspective for GenAI to empower education.

Keywords: Education Multi-agent; Student Portrait; Personalized Learning; Learner Model

1. 引言与相关工作

《深化新时代教育评价改革总体方案》明确指出需利用人工智能技术探索全要素评价，为学生提供精细的数字画像（Digital Profiling）。然而，当前的教育智能化实践存在局限性：其一，传统画像主要依赖行为或成绩等结构化数据，缺乏对学生认知潜能、情感体验及思维品质等隐性特征的挖掘（余明华等人，2021）；其二，大语言模型（LLM）虽具备强大的语义能力，但在面对长链条的教育推理任务时，单体智能体往往因注意力分散导致指令遵循能力下降（陈默等人，2025），生成的学业规划流于泛化。已有研究指出，多智能体学习系统能通过分布式协作分析学生表现并提供动态干预（Nazir et al., 2023），实现引导学习者有规划地推动学习进程的目标（吴永和等人，2024）。为此，为此，本研究依托 PISA 2022 数据集构建四维学生画像，并提出一种基于多智能体专家协同的教育智能架构，为大语言模型在教育领域的深度应用和因材施教的智能化教育提供新的理论视角和可行的技术路径。

2. 理论框架与系统架构

学生画像模型构建

学习者模型用于描述学习者认知、情感、学习行为等方面的特征，是学习分析与数据挖掘开展的基础（马志强和苏珊，2016）。本研究参照学习者模型理论与修订版布鲁姆分类学

(Krathwohl, 2002)，构建以知识、认知、情感、行为为四个核心维度的评估框架，既关注结果性的知识表现，又纳入反映思维过程的认知维度，以及影响学习成效的情感与行为维度。学生画像的整体构建流程如图 1 所示。数据来源于 OECD 发起的 PISA 2022 国际学生评估数据集，选取中国地区 5908 名学生为样本，从问卷与测试结果中筛选 67 个关键变量，构建四维评估体系：1) 知识维度以学科成绩 (PV) 为基准，结合数学问题解决频率 (ST276Q) 等变量评估情境化应用能力；2) 认知维度依托自我调节学习理论，选取任务坚持性 (ST307Q) 与多角度思考 (ST305Q) 测量认知韧性与批判性思维；3) 情感维度基于控制-价值理论，以数学焦虑 (ST297Q) 与学校归属感 (ST034Q) 识别情绪负荷与心理资本；4) 行为维度整合数字设备时长 (ST326Q) 与缺勤频率 (ST062Q)，作为内在心理状态的客观行为表征。

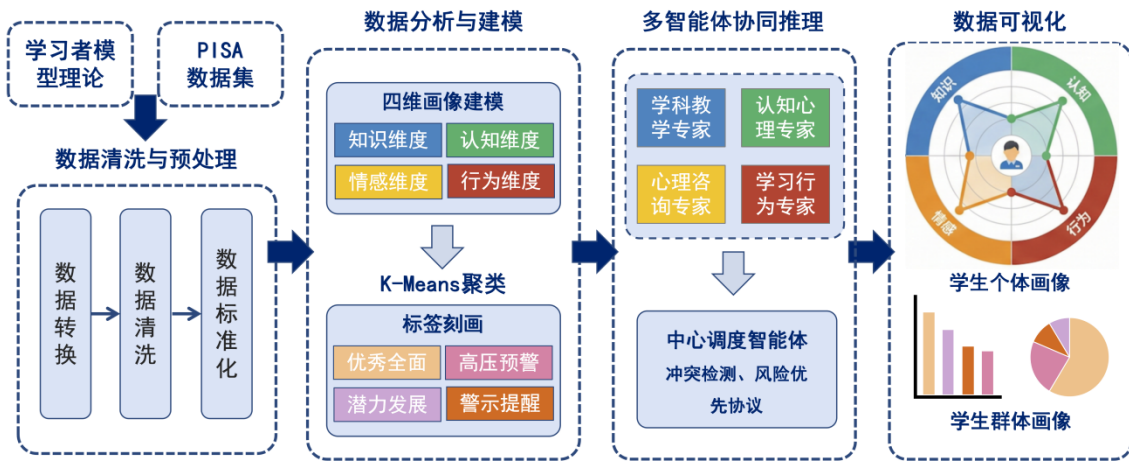


图 1 学生画像构建流程图

拟人化的多智能体教育协同架构

现有教育干预系统多停留于诊断层面，难以将数据分析结果转化为可执行的个性化规划。本系统构建了双回路协同机制，四位专家智能体承担诊断者与规划师两种角色，基于多维画像数据进行证据驱动的问题定位，并依托对应教育理论生成具体干预策略。

表 1 专家智能体诊断与规划设计

类	功能	核心理论
学科教学专家	诊断者：识别认知断层，分析成绩与问题解决矛盾，定位薄弱知识点。	ZPD 理论 布鲁姆认知分类学
	规划师：设计 ZPD 阶梯式学习题；推荐微课资源与拓展阅读清单。	
认知心理专家	诊断者：评估品质思维，坚持与分心的表现是策略问题还是意志力薄弱。	元认知理论 SRL 理论
	规划师：训练元认知策略，制定注意力训练计划；引导思维过程监控。	
心理咨询专家	诊断者：情绪风险反馈，识别高焦虑或低归属感等阻断性情感因素。	控制-价值理论 积极心理学
	规划师：提供情绪处方，生成引导话术，制定考前放松预案。	
学习行为专家	诊断者：投入度监测，关联缺勤与数字使用，识别低效时间管理。	行为主义强化 助推理论
	规划师：优化行为习惯，制定设备使用计划，设计助推激励机制。	

系统采用 CO-STAR (Context、Objective、Style、Tone、Audience、Response) 框架构建差异化提示词。核心设计为目标模块嵌入强制性行动指令，要求输出必须包含可操作步骤，从而实现从数据诊断到学习建议的闭环。针对 LLM 幻觉风险，提示词中明确限定模型推理范围，所有建议须锚定于输入的 PISA 字段数值，禁止生成无数据依据的推断，并结构化输出格式便于中心控制器解析与一致性核验。

当多位专家建议存在冲突时，中心调度智能体依据风险优先协议 (Risk-First Protocol) 执行仲裁。中心调度智能体首先通过本地预设规则判定风险等级，当情感或行为维度触发风险阈值时，系统自动将对应专家的建议赋予更高优先级；随后将风险等级与各专家输出共同注入提示词上下文，驱动中心智能体识别跨领域目标冲突，并在多目标优化框架下协调建议优先级。这一机制的本质是提示词驱动的动态判断，使系统能够适配教育场景高度异质的个体差异，而非以统一权重机械处理所有情境。最终中心调度智能体整合各专家输出，生成兼顾教育目标层次的规划，同时留存完整推理日志，确保系统决策过程的可解释性与可追溯性。

3.实证分析与系统有效性评价

基于 K-Means 聚类 (K=4)，系统将 5,908 名学生划分为四类群体。1) 优秀全面型 (Cluster A, 43.8%)：好成绩、低焦虑、良好行为习惯；2) 高压成绩型 (Cluster B, 6.7%)：典型隐性风险群体，情感维度标准化得分远超其他群体，高投入、高焦虑但中等成绩，极易被传统单维评价漏判；3) 潜力型 (Cluster C, 20.9%)：认知维度最高，成绩中等，具备元认知驱动的提升空间；4) 警示型 (Cluster D, 28.6%)：四维均处于劣势，需全方位补救干预。聚类结果验证了四维画像的分类效度：高压成绩型在单维成绩评价下完全不可见，而四维框架使其得以被系统性识别，这是本系统相对于传统评价的核心增量价值。针对该群体典型样本 (ID 34400004)，系统的多智能体并行推理与冲突解决日志如图 2 所示。

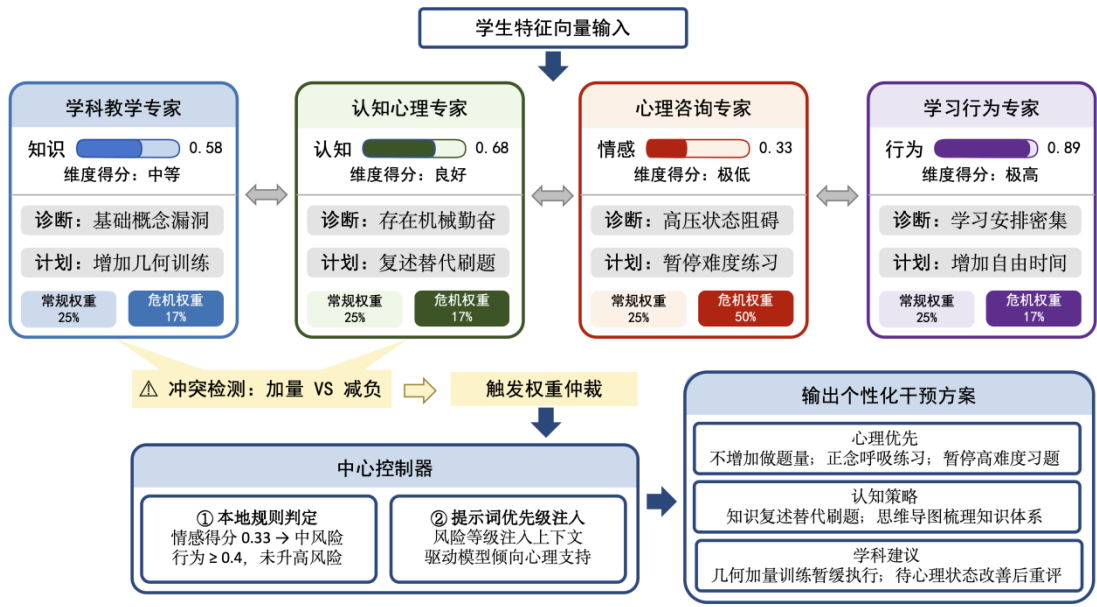


图 2 多智能体协同推理日志

在确认画像分类效度的基础上，本研究进一步评估系统生成规划的实际质量。从四类群体中各抽取若干学生，共取 20 名学生的画像数据及对应系统生成规划，邀请 3 位具有教育技术学背景的评价者对多智能体系统的输出进行双盲评分。评估量表依据 AERA 效度原则与 CASEL 社会情感学习框架制定，涵盖准确性、相关性、可行性与可操作性、个性化程度、教育合理性、语言流畅与专业性六个维度。

评估结果显示，六个维度均分均在 4.25-4.60/5.00，其中可操作性与语言专业性得分最高（均值 4.60），准确性与个性化程度次之（均值 4.50），相关性与教育合理性相对较低（均值 4.25），标准差介于 0.06–0.10 之间，评价者间一致性良好。质性反馈表明，生成建议能够有效契合 PISA 数据特征，且倾向于关注学生心理健康与长远发展，而非单纯追求分数提升。

多智能体架构相对于单体模型具有明显的差异化优势。针对 ID 34400004 样本，基线模型建议：引导其担任学科小组组长...将同理心转化为社交资本。事实上该生属于高压成绩型，通常伴随高敏感和社交退缩。让其担任领导职务可能会加剧其焦虑。基线模型仅看到了科学高分这一显性优势，未能发现高分背后的风险。相比之下，本系统在检测到高焦虑心理因素后，教育心理咨询师智能体明确建议：暂时不增加额外的社交负担，推翻了单纯基于成绩的学科专家的提议。系统原型提供了直观的验证，雷达图清晰展示了个人画像（蓝色区域）与班级平均（灰色边框）的差异。在实际测试中，专家反馈这种对比可视化能快速帮助其识别出偏科或心理高危学生，显著提升学情诊断的效率。



图 3 系统原型界面：画像雷达图与个性化规划展示

4.研究总结与展望

本研究提出并初步验证了一种多智能体专家协同的学生画像与学业规划架构。核心贡献主要体现在：第一，突破传统单维成绩的评价局限，基于学习者模型理论与修订版布鲁姆分类学构建了涵盖知识、认知、情感与行为的四维评估框架；第二，设计了角色分离的四专家协同架构，通过提示词框架与中心智能体调度将大语言模型从通用文本生成器设计为具备领域角色的教育专家智能体团队；第三，开发了可视化原型集成多维画像与个性化规划。PISA 问卷框架由 OECD 公开发布，教师只需组织学生完成 PISA 标准问卷即可提取所需的四维特征向量，驱动系统生成个体画像与干预规划，为系统向真实教学场景的规模化落地提供了现实基础。同时，本研究存在以下局限。当前验证样本规模有限，结论的外部效度有待更大规模研究检验；系统目前仅依赖结构化问卷数据，无法捕捉学生的纵向动态变化，亦难以反映课堂行为等细粒度个体信息。后续研究将引入追踪数据与多模态数据分析，以提升画像的动态适应性与干预建议的个性化程度。

参考文献

马志强,&苏珊.(2016).学习分析视域下的学习者模型研究脉络与进展.现代远距离教育,(4),44-50.

陈默,杨玉辉,杨清元,陈娟娟,&陈文智.(2025).智能体赋能高等教育变革:基于 DeepSeekR1 的范式重构与“浙大先生”实践探索.Modern Educational Technology,35(5).

吴永和,姜元昊,陈圆圆,&张文轩.(2024).大语言模型支持的多智能体:技术路径,教育应用与未来展望.开放教育研究,30(5),63-75.

余明华,张治,&祝智庭.(2021).基于学生画像的项目式学习评价指标体系研究.电化教育研究,42(3),89-95.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.

Nazir, M., Noraziah, A., & Rahmah, M. (2023). Students' performance prediction in higher education using multi-agent framework-based distributed data mining approach: A review. *International Journal of Virtual and Personal Learning Environments*, 13(1), 1-19.