

基于 AHP-模糊综合评价的编程能力动态监测与画像模型研究*

房楠 潘晟旻* 郝熙

昆明理工大学信息工程与自动化学院, 昆明 650500

摘要 针对目前的计算机程序设计课程中对学生代码评测仅以测试用例通过率为依据、评价维度单一、忽视学生的工程素养与阶段差距等问题, 本文在新工科人才培养背景下, 提出一种面向国家战略需求的多维编程能力评价框架, 引入层次分析法(AHP)与模糊综合评价(FCE)方法, 构建涵盖逻辑正确性、工程规范性、算法思维度与抗压健壮性的多维度编程能力动态监测与学生画像模型(Profiling Model), 并引入面向不同教学周期的自适应动态调权机制, 实现程序设计课程评价的多维化与个性化。典型样本验证表明, 该模型能够精准识别传统评测存在的虚高评价与虚低评价偏差, 例如将某传统满分样本的综合评价精准下修至 69.3 分, 将某传统仅 40 分样本的评价合理上修至 63.2 分, 评价结果更贴合学生代码的真实工程质量。

关键词 代码评测, 多维编程能力, 层次分析法, 模糊综合评价, 画像模型

Research on Dynamic Monitoring and Profiling Model of Programming Ability Based on AHP-Fuzzy Comprehensive Evaluation *

Nan FANG Shengmin PAN Xi Hao

School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
panshengmin@yeah.net

Abstract—To address the limitations of current computer programming courses—where student code assessment relies solely on test case pass rates, evaluation dimensions are single, and students' engineering competency and stage-based learning gaps are overlooked—this paper, under the context of New Engineering talent cultivation, proposes a multi-dimensional programming ability evaluation framework oriented toward national strategic needs. By integrating the Analytic Hierarchy Process (AHP) and Fuzzy Comprehensive Evaluation (FCE), a dynamic monitoring and student profiling model (Profiling Model) is constructed, covering logical correctness, engineering standards, algorithmic thinking, and robustness under stress. An adaptive dynamic weight adjustment mechanism oriented toward different teaching cycles is further introduced, achieving multi-dimensional and personalized evaluation of programming courses. Validation on typical samples shows that the model accurately identifies both inflated and deflated biases in traditional assessment—for instance, correcting a sample that traditionally received a perfect score down to 69.3 points, and adjusting one that traditionally scored only 40 points up to 63.2 points—so that the evaluation results better reflect students' actual code engineering quality.

Keywords—code assessment, multi-dimensional programming ability, analytic hierarchy process, fuzzy comprehensive evaluation, profiling model

1 引言

在加快高水平科技自立自强的战略背景下, 以大语言模型为代表的生成式人工智能在基础代码生成方

*基金资助: 全国高等学校计算机教育研究会教育研究项目“数字素养培育视域下青少年人工智能教育共同体建设路径研究与实践——数字人才培养计划”(CERACU2024R22); 云南省服务乡村振兴战略课程思政示范课培育项目“Python 语言程序设计”(2024KCSZ13)。

*通讯作者: 潘晟旻 panshengmin@yeah.net。

面性能优异, SonarQube 等 AI 代码评审工具也广泛应用于静态分析与漏洞检测^[1]。当机器已能高效完成基础编程任务时, 高校计算机人才培养模式亟需实现从“标准化生产”到“个性化成长”的嬗变^[2]。然而现有评价体系尚未跟上这一转变, 在课堂教学中, CodeRunner 等自动批改插件已在高校程序设计课程中广泛部署^[3], 但其核心评判逻辑仍高度依赖预设单元测试。在课外训练场景中, 国内洛谷、力扣等主流平台也普遍仅以用例通过率为唯一评分依据, 评价维

度单一，忽视了代码工程规范度、健壮性及学生思维演进，极易诱导学生养成面向测试用例编程的短视习惯，与新工科人才培养目标背道而驰。为此，本文引入层次分析法与模糊综合评价方法，以期程序设计课程评价机制升级提供理论支撑与算法路径。

2 编程能力与画像模型的多维度特征体系构建

以成果导向教育（OBE）理念为核心^[4]，遵循工程素养导向与多维特征正交原则，本文将编程能力目标集设为U，向下拆解为4个一级维度与9个二级指标，各指标定义与评价极向如表1所示。

表 1 多维画像特征量化体系表

一级特征维度	二级指标	指标说明	评价极向
逻辑正确性（ U_1 ）	基础用例通过率（ U_{11} ）	代码基本功能的实现程度	正向（+）
	边界用例覆盖率（ U_{12} ）	边界条件与异常输入的处理意识	正向（+）
工程规范性（ U_2 ）	圈复杂度（ U_{21} ）	代码嵌套深度与控制流复杂程度	负向（-）
	标识符命名语义度（ U_{22} ）	命名与上下文语义的相关程度	正向（+）
	有效注释覆盖率（ U_{23} ）	代码文档化素养的量化指标	正向（+）
算法思维度（ U_3 ）	时间复杂度（ U_{31} ）	实际时间开销与理论最优解的相对比值	负向（-）
	空间复杂度（ U_{32} ）	实际空间开销与理论最优解的相对比值	负向（-）
抗压健壮性（ U_4 ）	异常防御机制率（ U_{41} ）	防御性结构在代码中的覆盖密度	正向（+）
	内存泄漏风险指数（ U_{42} ）	动态内存未释放比例	负向（-）

（注：“+”表示数值越大素养越优，“-”表示数值越小素养越优。）

3 基于 AHP-模糊综合评价的动态画像模型设计

代码质量评价存在天然的灰色地带，传统固定权重的线性加权法难以处理工程规范度等具有显著模糊性的评价边界。AHP 与模糊综合评价的组合方法已在软件工程评价领域得到验证^[5]，本文将其迁移至代码评审质量的多维量化评价场景，并针对教学周期特点引入动态调权机制加以扩展。为此，本节首先通过 AHP 确立静态基准权重，继而利用模糊综合评价实现评价边界的柔性量化，并引入动态自适应调权矩阵匹配不同阶段教学目标。需指出的是，本文在方法论上进行了严格区隔，权重确立阶段采用经典 AHP 标度法以确保数学严谨性，综合评价阶段采用模糊隶属度函数解决硬性阈值带来的断崖式评分缺陷，规避了模糊 AHP 判断矩阵因缺乏数理逻辑基础而影响权重有效性的学术风险^[6]。

3. 1 基于 AHP 的静态基准权重确立

依据 Saaty 经典 1~9 标度法^[7]，邀请多位本校专业教师对各二级指标进行两两比较，构造判断矩阵 $A = (a_{ij})_{n \times n}$ ，其中 a_{ij} 表示指标 i 相对指标 j 的重要程度，且矩阵满足正互反性 $a_{ij} = 1/a_{ji}$ 。求解矩阵 A 的最大特征根 $\lambda_{\max}$ 及对应特征向量 W ： $AW = \lambda_{\max}W$ 。归一化后得到基准权重向量 $W = (w_1, w_2, \dots, w_n)^T$ 。采用一致性指标 CI 与一致性比率 CR 进行一致性检验： $CI = \frac{\lambda_{\max} - n}{n - 1}$ 、 $CR = \frac{CI}{RI}$ ，其中 RI 为平均随机一致性常量，如表 2 所示。当 $CR < 0.1$ 时，判断矩阵通过一致性检验，权重分配有效。这一基于专家评分与判断矩阵的 AHP 权重确定范式，同样是当前教育质量评价体系研究中的常见路径^[8]。

依据教学实践经验构造判断矩阵，经计算得 $\lambda_{\max} = 9.2028$ ， $CI = 0.0254$ ， $CR = 0.0175 < 0.1$ ，一致性满足要求。归一化后各二级指标基准权重如表 3 所示。

表 2 RI 数据表

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

表 3 指标对应权重表

指标	U_{11}	U_{12}	U_{21}	U_{22}	U_{23}	U_{31}	U_{32}	U_{41}	U_{42}
权重	0.2959	0.2034	0.0964	0.0552	0.0358	0.1430	0.0552	0.0900	0.0251

权重分配结果符合工程教育规律，逻辑正确性 $U_{11} + U_{12}$ 合计权重接近 50%，体现程序设计课程对功能正确性的基础性要求。算法思维度 $U_{31} + U_{32}$ 权重次之，反映对算法效率的重视。

3. 2 模糊测度矩阵与综合评价向量构建

设模糊评语集 $V = \{V_1, V_2, V_3, V_4\} = \{\text{优秀, 良好, 合格, 需重构}\}$ ，以隶属度函数将各指标特征值映射至对应等级概率。对于负向指标，等级临界值参数满足 $0 \leq a < b < c < d < e$ ，四个等级隶属度函数为：

优秀 V_1 $r_{i1}(x) = \begin{cases} 1, & x \leq a, \\ \frac{b-x}{b-a}, & a < x \leq b, \\ 0, & x > b. \end{cases} \quad (1)$

良好 V_2 $r_{i2}(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & b < x \leq c, \\ \frac{d-x}{d-c}, & c < x \leq d, \\ 0, & x > d. \end{cases} \quad (2)$

合格 V_3 $r_{i3}(x) = \begin{cases} 0, & x \leq b, \\ \frac{x-b}{c-b}, & b < x \leq c, \\ 1, & c < x \leq d, \\ \frac{e-x}{e-d}, & d < x \leq e, \\ 0, & x > e. \end{cases} \quad (3)$

需重构 V_4 $r_{i4}(x) = \begin{cases} 0, & x \leq d, \\ \frac{x-d}{e-d}, & d < x \leq e, \\ 1, & x > e. \end{cases} \quad (4)$

正向指标隶属度函数形式与负向指标对称， V_1 优秀对应高数值区间，隶属度函数图如图 1 所示。各指标等级临界值综合参考领域标准规范与专家经验确定，如表 4 所示。

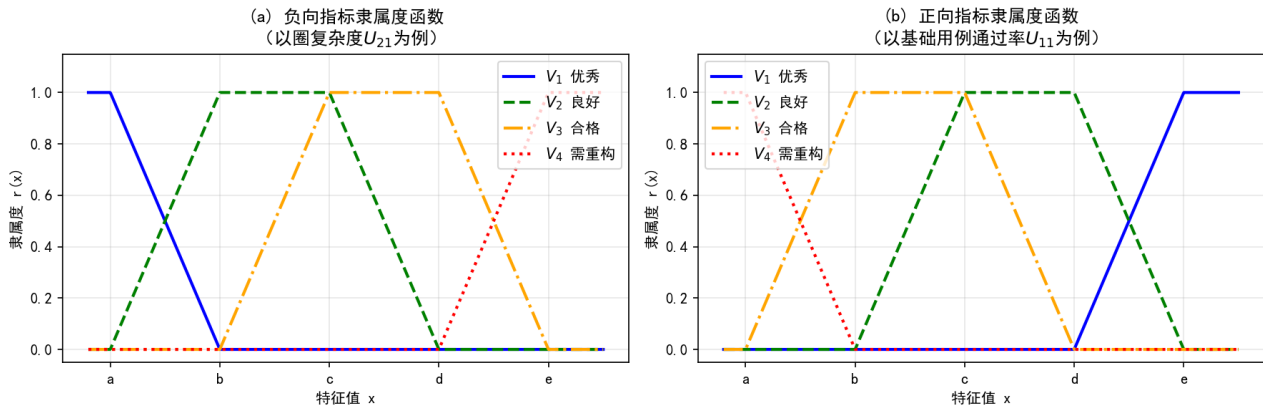


图 1 隶属度函数图

表 4 等级临界值参数表

指标	a	b	c	d	e
U_{11} 基础用例通过率 (%)	40	60	75	90	100
U_{12} 边界用例覆盖率 (%)	20	40	60	80	100
U_{21} 圈复杂度	5	10	20	30	50
U_{22} 标识符命名语义度 (0~1)	0.2	0.4	0.6	0.8	1.0
U_{23} 有效注释覆盖率 (%)	5	15	30	50	70
U_{31} 时间复杂度 (相对比值)	1.0	2.0	5.0	10.0	20.0
U_{32} 空间复杂度 (相对比值)	1.0	2.0	4.0	8.0	16.0
U_{41} 异常防御机制率 (%)	10	25	50	75	90
U_{42} 内存泄漏风险指数 (%)	0	5	15	30	50

将各指标特征值代入对应隶属度函数，并对各行进行归一化得标准模糊测度矩阵 $R_{n \times 4}$ ，将权重向量 W 与 $R_{n \times 4}$ 进行加权平均型模糊算子合成，得综合评价向

量 $B = W \circ R_{n \times 4} = (b_1, b_2, b_3, b_4)$ 。以某学生提交代码为例进行验证，合成后综合评价向量 $B =$

(0.3110,0.2039,0.2025,0.2827)，需重构等级隶属度达28.27%，验证了模型的有效性。

3.3 面向教学周期的动态自适应调权机制

传统 AHP 权重一旦确定便固定不变，难以适应高校计算机教学分阶段推进的客观规律。为此，本文设计时间驱动的动态自适应调权矩阵 $D(t)$ ，令 t 为当前教学阶段标识，针对各指标设计调节因子 $\alpha_i(t)$ ，例如工程规范性相关指标的调节因子随 t 单调递减，体现低年级阶段对规范养成的重点强调，而算法效率指标的调节因子随 t 单调递增，反映高年级阶段对算法效率要求的逐步提升。据此构建动态对角矩阵 $D(t) =$

$\text{diag}(\alpha_{11}(t), \alpha_{12}(t), \dots, \alpha_{42}(t))$ ，将静态权重向量 W 与动态调权矩阵 $D(t)$ 相乘并归一化，得到面向当前培养阶段的动态权重向量 $W(t) = \text{Normalize}(W \times D(t))$ ，相应地，动态综合评价向量更新为 $B(t) = W(t) \circ R$ 。

4 动态画像生成与智能预警仿真推演

4.1 典型学习者样本特征设计

依据一线教学中的常见编程缺陷构建四组典型代码样本：样本 A 唯结果导向，规范性严重不足。样本 B 规范高效但功能实现欠缺。样本 C 接近满分但算法严重臃肿。样本 D 为全维度参照基准。各指标特征值如表 5 所示。

表 5 四类典型样本特征值设定

指标	样本A	样本B	样本C	样本D
U_{11} 基础用例通过率（%）	100	40	95	100
U_{12} 边界用例覆盖率（%）	30	25	20	95
U_{21} 圈复杂度	28	6	35	5
U_{22} 命名语义度（0~1）	0.15	0.9	0.15	0.90
U_{23} 有效注释覆盖率（%）	0	62	0	60
U_{31} 时间复杂度（相对比值）	2.5	1.1	15.0	1.0
U_{32} 空间复杂度（相对比值）	2.5	1.1	12.0	1.0
U_{41} 异常防御机制率（%）	10	82	10	85
U_{42} 内存泄漏风险指数（%）	3	1	45	0

4.2 画像模型测度结果与传统评价的对比分析

将表 5 特征值代入模型，经解模糊化处理，将评语集 V 映射至对应百分制分数区间，即优秀[90,100]、良好[80,89]、合格[60,79]、需重构[0,59]，取各区间中值得到标准分列向量 $S = [95,85,70,30]^T$ ，计算最终期望得分 $E = B \cdot S$ 。与传统在线评测平台对比结果如表 6 所示。

对比结果表明，传统评测存在两类典型偏差，一是虚高评价，样本 A、C 传统平台给出满分与 95 分，而本文模型综合评价分别仅为 69.3 分与 54.9 分，精准识别出其代码的多维缺陷。二是虚低评价，样本 B 因功能实现不足传统平台仅给 40 分，本文模型综合评价为 63.2 分，客观肯定了其工程规范性与算法效率的优秀表现。基于标准模糊测度矩阵 $R_{n \times 4}$ 与标准分列向量 S 的矩阵运算绘制四类典型样本的多维画像雷达图如图 2 所示，为后续靶向干预提供客观数据坐标。

表 6 传统评测与本文模型测度对比

样本	传统在线评测平台得分	本文模型综合得分	本文模型检出的核心能力短板
样本A	100分(满分)	69.3分(合格)	U_{22} 、 U_{23} 、 U_{41} 评分偏低
样本B	40分(不及格)	63.2分(合格)	U_{11} 、 U_{12} 评分偏低
样本C	95分(满分)	54.9分(需重构)	U_{31} 、 U_{32} 评分偏低、 U_{42} 风险较高
样本D	100分(满分)	93.7分(优秀)	无明显短板

4.3 与雨课堂智能批改平台的横向比较

引入雨课堂智能批改平台进行横向比较，三种评价方式的系统对比如表 7 所示。雨课堂基于 LLM 驱动，在评价维度上较传统评测有所扩展，能对代码规范和

健壮性给出定性分项评分并生成逐行批注，但评价结果最终仍收敛为单一总分，无法输出各维度独立量化得分，也不支持跨教学周期的动态权重调整。但雨课堂设有代码原创性检测维度，可识别 AI 生成代码特

征，是本文模型尚未涵盖的方向，可作为后续研究的扩展。

表 7 三种评价方式横向对比

对比维度	传统在线评测	雨课堂智能批改	本文模型
评价输出形式	单一通过率得分	4项分项得分+总分+逐行批注	9维画像向量+雷达图+期望得分
功能正确性评价	✓（黑盒通过率）	✓（功能实现项）	✓（ U_{11} ）
边界覆盖评价	×	部分（健壮性项定性描述）	✓（ U_{12} 量化）
代码规范量化	×	部分（结构定性评分）	✓（ $U_{21}/U_{22}/U_{23}$ 量化）
算法时空效率	间接（运行时间）	×	✓（ U_{31}/U_{32} 量化）
内存安全评价	×	×	✓（ U_{42} 量化）
AI生成代码检测	×	✓（代码原创项）	×
个性化动态调权	×	×	✓（ $D(t)$ 机制）
评价边界处理	硬性阈值	硬性阈值	模糊隶属度函数
多维画像输出	×	×	✓（雷达图可视化）

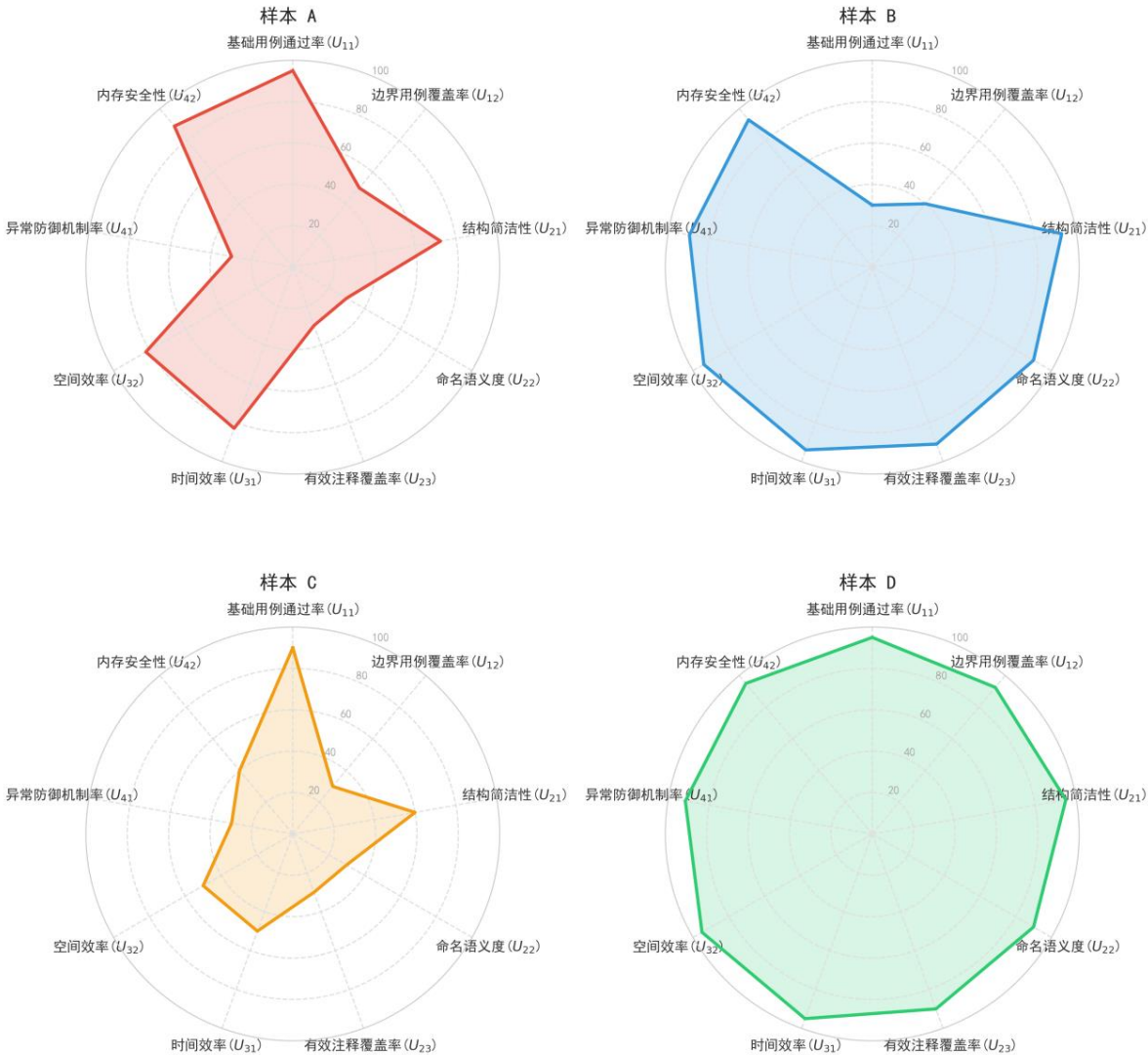


图 2 多维画像雷达图

（注：各维度得分越高表示编程能力表现越优，负向指标经模糊映射后正向化，高分对应低圈复杂度、高时间效率、高空间效率、低内存泄漏风险。）

5 结束语

本文构建了涵盖逻辑正确性、工程规范性、算法思维度与抗压健壮性的多维编程能力画像模型,利用AHP与模糊综合评价将代码底层特征数据柔性映射为综合画像等级,并通过动态自适应调权矩阵赋予模型跨教学周期的动态调整能力。典型样本验证表明,该模型能有效识别传统评测的虚高与虚低偏差。作为探索性研究,模型实际教学效果有待真实平台对照实验验证。在此基础上,后续工作将沿两条路径推进:一是构建AHP算法测度与LLM启发式反馈协同干预框架,实现短板特征的精准定位与个性化预警。二是引入数据驱动的权重动态修正机制、融合RAG技术防范推理幻觉,并依托相关智慧校园平台开展长效跟踪实证研究。

参考文献

- [1] 曾新励. 基于 SonarQube 平台提升企业软件项目质量的应用研究[J]. 电脑知识与技术, 2019, 15(11): 78-80.
- [2] 罗蓉, 张安富, 王剑. AI 时代高校教育教学数智化变革的理性思考[J]. 高教发展与评估, 2026, 42(02): 1-9+181.
- [3] 徐兴, 张智刚, 孙道宗, 等. 基于 Coderunner 的程序设计作业在线评测与实践[J]. 创新创业理论研究与实践, 2022, 5(07): 26-28.
- [4] 肖宁, 高敏, 李毅, 等. 新工科背景下数据挖掘课程思政建设探索[J]. 计算机教育, 2026, (02): 75-81.
- [5] 许媛媛. 基于 AHP 与模糊综合评价模型的软件过程绩效评价方法[J]. 软件, 2025, 46(03): 105-109.
- [6] 朱克毓, 杨善林. 关于 Saaty 对模糊逻辑不适用于 AHP 观点的评述[J]. 系统工程理论与实践, 2014, 34(01): 197-206.
- [7] Saaty R W. The analytic hierarchy process—what it is and how it is used[J]. Mathematical modelling, 1987, 9(3-5): 161-176.
- [8] 姚怡, 梁微, 孙宇. 基于 CIPP+AHP 的在线开放课程质量评价体系构建[C]//第 31 届全国计算机新科技与教育学术会议. 全国高校计算机教育研究会, 2023.