Computer Science and Application 计算机科学与应用, 2026, 16(5), 183-195
Published Online May 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.165175

Hans汉斯

基于多智能体协同的动态教学优化系统

阮子豪，冯婧琳，李　晶

北京工商大学计算机与人工智能学院，北京

收稿日期：2026年4月18日；录用日期：2026年5月14日；发布日期：2026年5月26日

摘　要

针对传统智能教学系统缺乏动态适应性、忽视学生心理状态以及师生交互不足等问题，文章提出了一种基于多智能体强化学习的个性化教学系统。该系统通过学生智能体和教师智能体的协同优化，实现教学策略的动态调整。学生智能体基于Q-Learning算法，综合考虑掌握度、动机、疲劳度等多维状态构建智能体状态与动作空间，自主选择最优学习动作；教师智能体融合启发式规则与强化学习，根据学生状态制定个性化的教学节奏、难度和策略；评估智能体负责对学习效果进行深度诊断与可解释反馈。该系统能有效提升学生的知识点掌握度，提高教师教学质量。同时，系统集成DeepSeek大语言模型，实现个性化学习报告的生成，为用户提供沉浸式的交互体验。研究为智能教学系统的设计提供了新的思路，具有良好的理论价值和应用前景。

关键词

强化学习，多智能体系统，个性化教学，Q-Learning，智能教学系统

Dynamic Teaching Optimization System Based on Multi-Agent Collaboration

Zihao Ruan, Jinglin Feng, Jing Li

School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing

Received: April 18, 2026; accepted: May 14, 2026; published: May 26, 2026

Abstract

Aiming at the problems of traditional intelligent tutoring systems, such as a lack of dynamic adaptability, neglect of students' psychological states, and insufficient teacher-student interaction, this paper proposes a personalized teaching system based on multi-agent reinforcement learning. The system realizes the dynamic adjustment of teaching strategies through the collaborative optimization of student agents and teacher agents. The student agent, based on the Q-Learning algorithm,

文章引用: 阮子豪, 冯婧琳, 李晶. 基于多智能体协同的动态教学优化系统[J]. 计算机科学与应用, 2026, 16(5): 183-195. DOI: 10.12677/csa.2026.165175

constructs the agent state and action space by comprehensively considering multidimensional states including knowledge mastery, motivation, and fatigue level, and independently selects the optimal learning action; the teacher agent integrates heuristic rules and reinforcement learning to formulate personalized teaching rhythm, difficulty level, and strategies according to students' states; the evaluation agent is responsible for in-depth diagnosis of learning effects and interpretable feedback. The system can effectively improve students' knowledge mastery and enhance the quality of teachers' teaching. Meanwhile, the system integrates the DeepSeek large language model to realize the generation of personalized learning reports and provide users with an immersive, interactive experience. The work of this paper provides a new idea for the design of intelligent tutoring systems, with good theoretical value and application prospects.

Keywords

Reinforcement Learning, Multi-Agent System, Personalized Teaching, Q-Learning, Intelligent Tutoring System

1. 引言

随着“人工智能 + 教育”上升为国家战略，国家政策明确指向利用智能技术推动人才培养模式改革，构建以学习者为中心的新型教育体系[1]-[3]。在此背景下，在线学习平台与智慧课堂等数字化教育形态蓬勃发展，积累了海量教学过程数据，为精准化、个性化的智能教育提供了数据基础。

然而，当前教育技术应用仍面临诸多挑战。一方面，现有研究多采用单智能体交互[4]，未能发挥多智能体协同演化的系统效能，反馈内容单一，难以通过协作机制持续提升教学效果[5]。另一方面，学生的动机、疲劳度等认知与情感状态常被忽略，导致推荐策略与学生真实接受能力脱节。因此，构建能够模拟教学相长过程、实现教学策略与学习行为双向协同优化的系统，成为关键难题。

多智能体系统为解决上述挑战提供了新思路。将强化学习引入多智能体系统，可使智能体在与环境动态交互中通过试错不断优化策略[6]。应用于教育场景，可将教师与学生的决策分别建模为独立智能体：教师智能体以最大化班级整体学习效果为目标，基于班级整体知识掌握状态动态调整教学节奏、难度与策略；学生智能体则在教师引导下，依据自身知识水平、学习偏好与认知状态选择最优学习动作，以最大化个人学习收益[7]。二者在同一教学环境中相互交互、共同进化，形成能够模拟并优化真实教学过程的自适应系统，为实现“规模化”与“个性化”统一的教育范式提供了可行的技术路径。

2. 多智能体协同学习机制

本研究旨在设计并实现一个基于多智能体强化学习的动态教学优化系统，设计了包含学生智能体、教师智能体和评估智能体的多智能体协同学习框架。三个智能体分别承担学习执行、教学引导和效果评价的职能，通过信息共享与策略协同，共同优化学生的学习过程。

2.1. 系统框架与角色定义

参照[8]提出的多智能体教学思路，多智能体系统整体架构如图 1 所示。学生智能体直接与教师智能体交互，执行具体的学习动作并更新自身状态；教师智能体观测学生状态，制定教学策略并推荐学习动

作，引导学生探索；评估智能体在每轮学习结束后，综合分析学生的学习轨迹与最终状态，生成诊断报告与改进建议，反馈给教师和学生以优化后续学习。

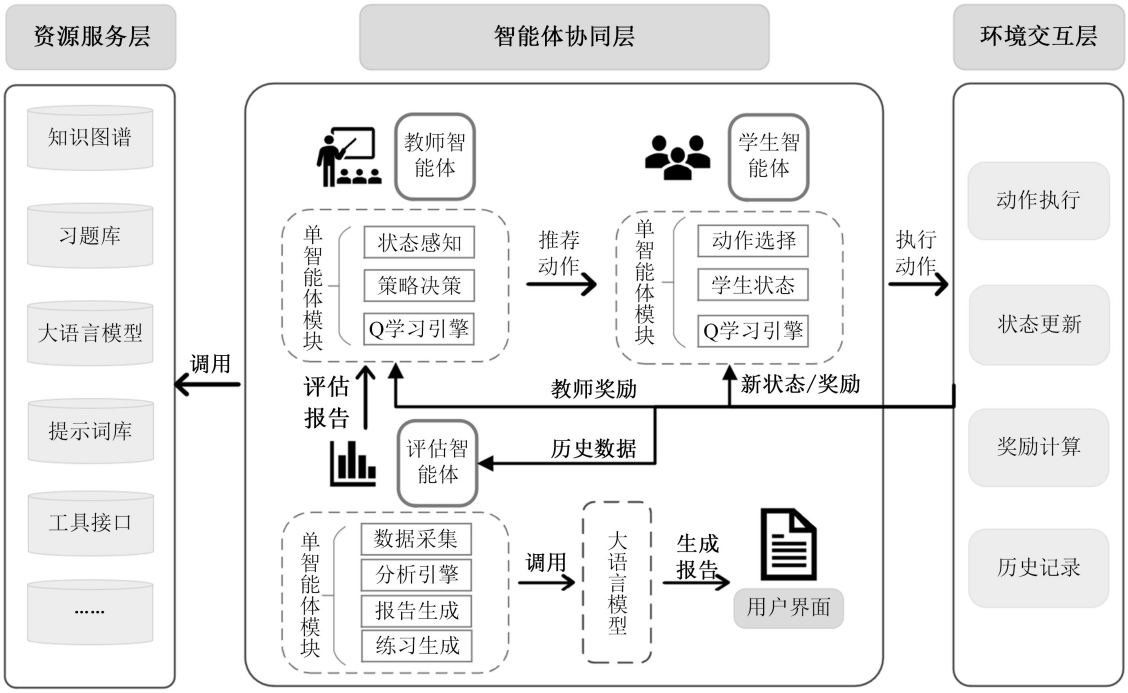

Figure 1. Technical architecture of multi-agent base on teaching optimization system
图 1. 基于多智能体的教学优化系统技术架构

学生智能体的核心是状态空间 S_S、动作空间 A_S 和基于 Q-Learning 的学习机制。学生负责执行学习动作，并根据即时奖励更新自身策略；教师智能体其状态空间 S_T、动作空间 A_T 均基于学生状态构建，通过强化学习优化教学策略，并通过引导机制影响学生动作选择；评估智能体其输入为学生的学习历史 H_S (包括所有动作序列、状态变化及奖励)和最终状态 s_F，输出为综合性评估报告 R。评估报告包含知识点掌握分析、学习效率评价、薄弱环节识别以及个性化学习建议。评估智能体不直接参与在线决策，而是作为后验分析工具，为教师和学生的策略调整提供全局视角。

2.2. 协同工作机制

文献[9]提出了一种完全合作的智能体协作模式，本研究中在参考后让三个智能体采取“感知–决策–执行–评估”的循环协同工作，具体流程如下：

1) 在每一步 t，教师智能体观测当前学生状态 s_t^S，通过离散化得到自身状态 s_t^T，并依据其策略选择教学动作 a_t^T，生成推荐学生动作 a^{rec}。

2) 学生智能体接收推荐动作，将其作为 Q 值修正的引导，然后依据 ε-贪心策略选择实际执行的动作 a_t^S。

3) 学生执行动作 a_t^S，环境根据学生模型更新状态至 s_{t+1}^S，并返回即时奖励 r_t^S。学生据此更新自身 Q 表。

4) 教师智能体根据观察到的掌握度变化 $\Delta m_{i,t}$ 计算自身奖励 r_t^T，并更新自身 Q 表。

5) 当学习过程结束(达到最大步数或满足终止条件)，评估智能体读取完整学习历史 H_S 和最终状态 s_F，调用大模型生成评估报告 R。

该协同机制能够根据学生状态动态调整教学策略，实现个性化自适应，同时兼顾探索效率与学习自主性，显著提升学生的学习效率与最终掌握度，尤其在处理多知识点、长周期学习任务时表现出良好的适应性和鲁棒性。

3. 智能体的设计

3.1. 学生智能体的设计

3.1.1. 状态空间

学生智能体状态由固有属性、动态状态与历史记录构成，第 t 步状态为：

$$s_t^S = \left(\Theta, \boldsymbol{M}_t, H_S\right) \in S_S \tag{1}$$

1) 固有学习属性

固有属性仅在初始化时确定，不随学习过程改变：

$$\Theta = \left(id, \theta, \mathcal{K}, \mathcal{I}\right) \tag{2}$$

其中，id 是学生的唯一标识，θ 服从 $U(0.9, 1.3)$的均匀分布，作为学生的基础学习能力，$\mathcal{K}$ 为知识点的集合，$\mathcal{I}: k_i \mapsto i$ 为知识点索引映射，用于向量化计算。

2) 动态学习状态

动态学习状态是学生当前学习状态的核心表征，随动作执行实时更新，动态学习状态的组成逻辑如式(3)所示：

$$\boldsymbol{M}_t = \left(\boldsymbol{m}_t, \mu_t, \phi_t\right) \tag{3}$$

上式中 $\boldsymbol{m}_t$ 是知识点掌握度向量，$\boldsymbol{m}_t = \left[m_{1,t}, m_{2,t}, \cdots, m_{n,t}\right]^{\mathrm{T}}$，$m_{i,t} \in [0,1]$，当 $m_{i,t} = 0$ 时表示学生对该知识点完全未掌握，如果 $m_{i,t} = 1$，则表示学生对该知识点 i 完全掌握，μ_t 是学生的学习动机($\mu_t \in [0.5, 1.5]$)，随动作的执行更新按照 $\mu_t = \mu_{t-1} + \Delta\mu\left(a_t^S\right)$ 逻辑更新，$\Delta\mu\left(a_t^S\right)$ *为动作* a_t^S *对应的动机变化量*；$\phi_t \in [0, 1.0]$ 为学生的学习疲劳度，初始值 $\phi_0 = 0$，随动作执行按照 $\phi_t = \phi_{t-1} + \Delta\phi\left(a_t^S\right)$ 逻辑累积，$\Delta\phi\left(a_t^S\right)$ 为动作 a_t^S 对应的疲劳度变化量。

3) 历史行为记录

历史行为记录 H_S 反映学习过程特征，为策略优化提供历史依据：

$$H_S = \left(T_t, S_t, F_t, \mathcal{L}_t\right) \tag{4}$$

T_t 是学生的总学习耗时(单位：分钟)，$\mathcal{L}_t$ 记录动作、知识点、掌握度变化与奖励，S_t 和 F_t 分别是学生答题过程中的连续成功次数和连续失败次数，相关定义如下：

$$S_t = \begin{cases} S_{t-1} + 1, & \Delta m_{i,t} > 0.05 \\ 0, & \text{其他} \end{cases} \tag{5}$$

$$F_t = \begin{cases} F_{t-1} + 1, & \Delta m_{i,x} < 0 \\ 0, & \text{其他} \end{cases} \tag{6}$$

4) 状态转移规则

执行动作 $a_t^S \in A_S$ 后，状态从 s_t^S 转移至 s_{t+1}^S；状态转移前后的固有属性 Θ 保持不变，动态状态 $\boldsymbol{M}_{t+1}$ 与历史记录 H_{t+1} 由动作效果更新。

$$s_{t+1}^S = \left(\Theta, \boldsymbol{M}_{t+1}, H_{t+1}\right) \tag{7}$$

3.1.2. 动作空间

动作空间涵盖 8 类典型学习行为，相关定义以及动作执行后对状态的影响效果如表 1 所示：

$$A_S = \{a_1, a_2, \cdots, a_8\} = \{\text{做练习题}, \text{看讲解视频}, \text{请教老师}, \text{小组讨论}, \text{复习错题}, \text{做挑战题}, \text{知识梳理}, \text{切换知识点}\} \tag{8}$$

Table 1. Action space table
表 1. 动作空间表

动作编号	动作类型	掌握度增益	耗时(分钟)	疲劳度增加	动机变化
a_1	做练习题	0.12	10	0.15	+0.05
a_2	看讲解视频	0.08	15	0.08	+0.03
a_3	请教老师	0.15	8	0.10	+0.08
a_4	小组讨论	0.10	12	0.10	+0.10
a_5	复习错题	0.14	8	0.12	+0.06
a_6	做挑战题	0.18	12	0.20	+0.12
a_7	知识梳理	0.08	10	0.05	+0.04
a_8	切换知识点	0.03	5	0.03	+0.02

动作执行后，知识点 k_i 的实际掌握度提升值由基础值结合学生状态修正得到：

$$\Delta m_{i,t} = g_j \cdot \theta \cdot \mu_t \cdot (1 - 0.5 \cdot \phi_t) \cdot (1 - 0.5 \cdot m_{i,t}) \cdot \eta \tag{9}$$

其中 $\eta \in [0.9, 1.1]$ 为教师教学策略影响因子。

3.1.3. Q-Learning 算法实现

1) 奖励函数

在考量了文献[10]中提出的 Q 学习算法后，本研究以学生知识点掌握度的有效提升为核心评价标准，综合考虑学习效率、连续学习表现等关键因素，通过设置不同权重系数构建多维度奖励函数，既引导学生智能体追求知识掌握度的提升，又约束其避免无效的耗时行为，同时对连续的学习成功与失败做出差异化奖惩，设计以下的奖励函数：

$$r_t^S = w_1 \cdot \Delta m_{i,t} + w_2 \cdot \text{bonus}(S_t) + w_3 \cdot \text{penalty}(F_t) - w_4 \cdot \frac{t(a_t^S)}{T_{\text{avg}}} \tag{10}$$

式中，$w_1 \cdot \Delta m_{i,t}$ 为知识点掌握度提升带来的基础奖励，是奖励函数的核心组成；$w_2 \cdot \text{bonus}(S_t)$、$w_3 \cdot \text{penalty}(F_t)$、$w_4 \cdot \frac{t(a_t^S)}{T_{\text{avg}}}$ 分别为连续成功奖励、连续失败惩罚与耗时惩罚。当学生出现连续的掌握度有效提升时，给予额外奖励以强化正向学习行为，反之当学生出现连续的学习无效甚至掌握度下降时，给予负向奖励以规避低效学习动作的重复选择；与此同时，将单次学习动作的耗时与平均耗时作对比，对超出平均耗时的行为进行惩罚。

2) Q 值更新目标方程

学生智能体的 Q 值定义为 $Q^S(s_t^S, a_t^S)$，其中 s_t^S 为学生当前的学习状态，$a_t^S \in A_S$ 为学生在该状态下执行的学习动作。Q 值表征了在状态 s_t^S 下执行动作 a_t^S 所能获得的长期累积奖励期望，是学生智能体动作选择的核心依据。对算法[11]改进后得到学生智能体的 Q 值更新核心公式如下：

$$Q^S\left(s_t^S,a_t^S\right)\leftarrow Q^S\left(s_t^S,a_t^S\right)+\alpha_t\left[r_t^S+\gamma^S\max_{a'\in A_S}Q\left(s_{t+1}^S,a'\right)-Q^S\left(s_t^S,a_t^S\right)\right] \tag{11}$$

其中，$\alpha_t=0.15\times\left(1+\left|r_t^S\right|\times 0.1\right)$ 为动态学习率，区别于固定学习率，该设计使奖励绝对值越大，学习率越高，能够加快智能体对有效或无效学习动作的学习速度，提升策略优化效率；$\gamma^S=0.9$ 为折扣因子，该取值使学生智能体既关注当前的学习效果，又不忽视长期的知识积累；$\max_{a'\in A_S}Q\left(s_{t+1}^S,a'\right)$ 为下一状态 s_{t+1}^S 下所有可执行动作的最大 Q 值，表征了学生智能体在后续学习中能够获得的最优奖励期望；$\delta_t=\gamma^S\max_{a'\in A_S}Q\left(s_{t+1}^S,a'\right)$ 为时序差分误差，当 $\delta_t>0$ 时表明当前动作的实际收益高于预期，将上调对应 Q 值，当 $\delta_t<0$ 时，表明当前动作的实际收益低于预期，将下调对应 Q 值。

3) 动作选择逻辑

学生智能体采用 ε-贪心策略实现“探索–利用”平衡，该策略能够使智能体在学习初期以较高概率探索不同的学习动作，挖掘潜在的最优策略，在学习后期以较高概率利用已学习到的最优动作，其具体的动作选择规则为：

$$a_t^S=\begin{cases}\arg\max_{a_t^S\in A_S}Q^S\left(s_t^S,a_t^S\right), & 1-\varepsilon_t\\ \text{random}\left(a_t^S\in A_S\right), & \varepsilon_t\end{cases} \tag{12}$$

其中，探索率 ε_t 采用指数衰减策略：

$$\varepsilon_t=\max\left(\varepsilon_{\min},\varepsilon_{t-1}\times\lambda\right) \tag{13}$$

初始探索率 $\varepsilon_0=0.4$，衰减系数 $\lambda=0.99$，探索率最小值 $\varepsilon_{\min}=0.05$，通过该设计实现智能体“前期多探索、后期多利用”的学习策略，既保证了策略探索的全面性，又兼顾了后期学习的高效性。

通过上述奖励函数，Q 值更新规则与动作选择逻辑的设计，Q-Learning 算法在学生智能体中实现了有效落地，使学生智能体能够在与教学环境及教师智能体的交互中，持续优化学习动作选择策略，实现学习效果的逐步提升。学生智能体具体学习流程如图 2 所示。

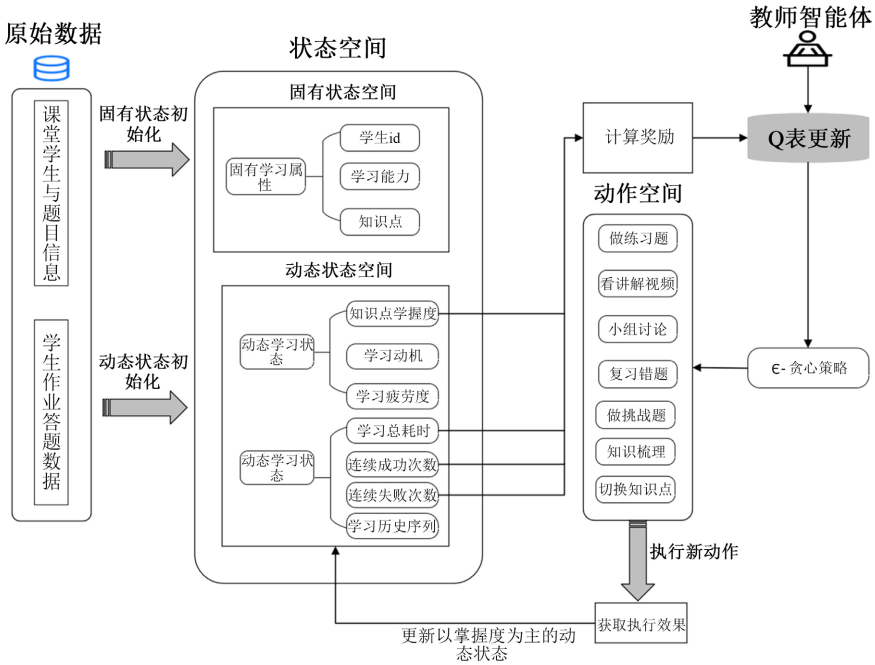

Figure 2. Schematic diagram of learning logic for student agent
图 2. 学生智能体学习逻辑示意图

3.2. 教师智能体设计

教师智能体在个性化学习系统中负责制定教学策略，其核心任务是根据学生当前学习状态动态调整教学节奏、难度与策略，并向学生推荐最优学习动作，以最大化学习效果。教师智能体的决策作用于教学组织层面，通过与学生的交互实现策略的自适应优化[12]。

3.2.1. 教师状态空间

教师智能体的决策依赖于对学生状态的全面感知。为降低状态空间的维度并提高学习效率，将学生状态中的连续变量进行离散化处理，构建教师智能体的状态表示。定义教师状态 s_t^T 为学生知识点掌握度水平与学习动机水平的组合：

$$s_t^T = \left(\ell_t^{\text{mastery}}, \ell_t^{\text{motivation}}\right) \in S_T \tag{14}$$

其中，ℓ_t^{mastery} 为当前学习知识点 k_i 的掌握度等级，根据掌握度 $m_{i,t}$ 划分为五个级别：

$$\ell_t^{\text{mastery}} = \lfloor 5 \cdot m_{i,t} \rfloor \in \{0,1,2,3,4\} \tag{15}$$

$\ell_t^{\text{motivation}}$ 为学习动机等级，依据动机值 μ_t 划分为三个区间：

$$\ell_t^{\text{motivation}} = \begin{cases} 0, & \mu_t < 0.8 \\ 1, & 0.8 \le \mu_t \le 1.2 \\ 2, & \mu_t > 1.2 \end{cases} \tag{16}$$

实际应用中也可根据具体场景引入疲劳度等变量，但为简化模型并保证收敛速度，本文以掌握度与动机为主要依据。教师状态空间的大小为 $|\mathcal{K}| \times 5 \times 3$，其中 $|\mathcal{K}|$ 为知识点总数。

3.2.2. 教师动作空间

教师智能体的动作定义为教学策略的三元组，涵盖教学节奏、教学难度和教学策略类型，每个分量均有明确的语义并对应学生动作空间中的推荐动作：

$$a_t^{\text{T}} = \left(r_t, d_t, p_t\right) \in \mathcal{A}_T \tag{17}$$

其中，教学节奏 $r_t \in$ {慢速，中速，快速}，教学难度 $d_t \in$ {简单，中等，困难}，教学策略 $p_t \in$ {巩固基础，拓展提高，复习弱点，启发思考，激励鼓励}，每种策略映射至一个推荐的学生动作 a^{rec}，映射关系如表 2 所示。

Table 2. Teaching strategy-action mapping table
表 2. 教学策略－动作映射表

教学策略	推荐学生动作
巩固基础	做练习题
拓展提高	做挑战题
复习弱点	复习错题
启发思考	小组讨论
激励鼓励	看讲解视频

3.2.3. 策略选择机制

教师智能体采用 ε-贪心策略[13]平衡探索与利用，并融入基于教学常识的启发式规则以加速初期收敛，其中，启发式特征是指一些经验性的但通常有效的规则或条件，用于简化问题的求解过程或提高效

率[14]。在状态 s_t^T 下，教师以概率 $1-\varepsilon_t$ 选择当前 Q 值最大的动作，以概率 ε_t 进行探索。探索时依据启发式规则确定各动作分量的概率分布，规则如下：

- 若 $m_{i,t}<0.3$ 或 $\phi_t>0.7$，则偏好“慢速”节奏；
- 若 $m_{i,t}>0.7$ 且 $\phi_t<0.3$，则偏好“快速”节奏；
- 若 $m_{i,t}<0.3$，则偏好“简单”难度与“复习弱点”策略；
- 若 $m_{i,t}>0.7$ 且 $\mu_t>1.2$，则偏好“困难”难度与“拓展提高”策略；
- 若 $\phi_t>0.6$，则偏好“激励鼓励”策略以缓解疲劳。

将上述规则转化为各动作分量的概率分布，再组合为完整动作的概率。设探索时选择节奏 r、难度 d、策略 p 的概率分别为 $P_R(r)$、$P_D(d)$、$P_P(p)$，则联合概率为 $P(r,d,p)=P_R(r)P_D(d)P_P(p)$，从该分布中采样得到探索动作。

3.2.4. 奖励函数与 Q 值更新

教师奖励反映教学策略的有效性，以学生执行推荐动作后的掌握度提升为核心，同时兼顾学习效率：

$$r_t^T=\alpha_1\cdot\Delta m_{i,t}+\alpha_2\cdot\mathbb{I}\left(\Delta m_{i,t}>0.05\right)-\alpha_3\cdot\frac{t\left(a_t^s\right)}{T_{\text{avg}}} \tag{18}$$

其中，$\Delta m_{i,t}$ 为掌握度实际增益，$\mathbb{I}(\cdot)$ 为指示函数，$t\left(a_t^s\right)$ 为动作耗时，T_{avg} 为平均耗时，α_1、α_2、α_3 为权重系数(取 $\alpha_1=5$、$\alpha_2=2$、$\alpha_3=1$)。该奖励函数兼顾了掌握度提升的幅度与效率，鼓励教师选择能带来显著进步且耗时合理的策略。

教师 Q 值采用标准 Q-Learning 更新规则[15]，使用教师自身的奖励和下一状态的最大 Q 值：

$$Q^T\left(s_t^T,a_t^T\right)\leftarrow Q^T\left(s_t^T,a_t^T\right)+\beta_t\left[r_t^T+\gamma^T\max_{a'\in\mathcal{A}_T}Q^T\left(s_{t+1}^T,a'\right)-Q^T\left(s_t^T,a_t^T\right)\right] \tag{19}$$

其中，$\beta_t=0.15\times\left(1+0.1\left|r_t^T\right|\right)$ 为动态学习率，$\gamma^T=0.9$ 为折扣因子。s_{t+1}^T 由学生执行动作后的新状态经离散化得到。

3.2.5. 教师－学生交互设计

教师智能体与学生的协同体现在教师输出的教学策略直接影响学生的动作选择。教师选择的策略 p_t 映射出推荐动作 a^{rec}，学生智能体在选择动作时对该推荐动作的 Q 值施加一个正向偏置：

$$\tilde{Q}^S\left(s_t,a\right)=Q^S\left(s_t,a\right)+\lambda\cdot\mathbb{I}\left(a=a^{rec}\right) \tag{20}$$

其中，$\lambda=0.5$ 为引导强度系数，$\tilde{Q}^S$ 为学生实际用于决策的修正 Q 值。学生在此基础上采用 ε-贪心策略选择最终动作 a_t^S。该机制既保留了学生的自主探索能力，又使教师策略能够有效引导学习过程。

学生执行动作后，环境反馈新的状态和即时奖励，学生更新自身 Q 表，同时教师根据观察到的效果计算奖励 r_t^T 并更新自身 Q 表，形成一个完整的交互闭环。

3.3. 评估智能体设计

作为系统的第三个关键组件，评估智能体负责对学习效果进行深度诊断与可解释反馈。该智能体以大语言模型 DeepSeek-Chat 为技术基座，通过结构化提示词将强化学习训练过程中产生的数值化状态序列转化为符合教育场景的自然语言报告，实现“教学－学习－评估”闭环中的诊断与反馈环节。

3.3.1. 总体架构与处理流程

评估智能体的输入由三部分构成：学生智能体的最终状态，包括各知识点掌握度、学习动机、疲劳度、总学习时长；训练过程摘要，如最近若干步的学习动作序列及掌握度变化；班级层面的统计信息如

各知识点班级平均掌握度。原始数据经预处理模块转换为格式化文本后，填入预设的提示词模板，最终调用大语言模型生成结构化的个性化学习报告。整体处理流程如图 3 所示。

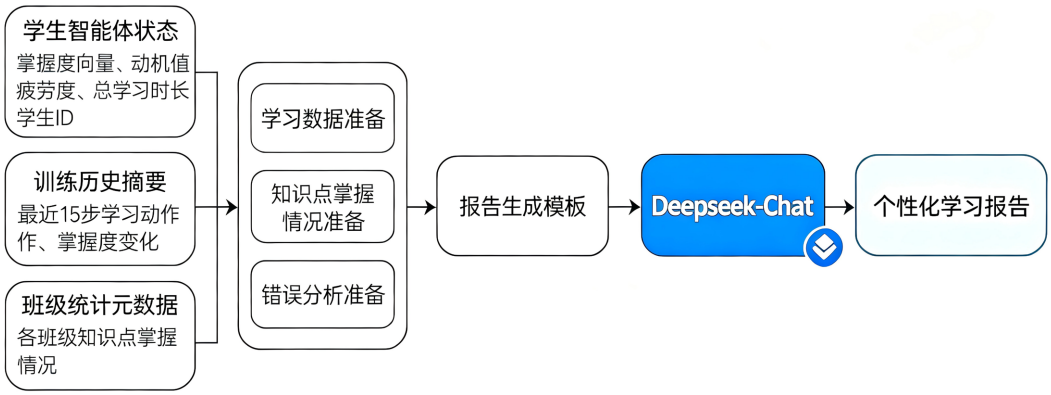

Figure 3. Model report generation and modular data processing flowchart
图 3. 模型报告生成及模块数据处理流程

3.3.2. 提示词模板设计

为引导大模型生成符合教育场景要求的报告内容，系统设计了两级提示词模板。主模板规定了报告的标准结构，包含五个核心模块：

1) 学习效果总体评估：基于学生整体掌握度、学习时长等宏观指标，给出概括性评价。

2) 薄弱知识点分析：针对掌握度低于预设阈值(60%)的知识点，结合班级平均水平和学生历史表现，分析其掌握不足的可能原因。

3) 个性化学习建议：提供 3~5 条具体可行的改进建议，涵盖学习方法、时间分配、重点攻克方向等维度。

4) 推荐学习策略：基于学生当前的认知状态(动机、疲劳度)和知识掌握分布，推荐最适合该学生的教学节奏、难度和策略组合。

5) 针对性练习题：为掌握度最低的三个知识点各生成一道练习题，包含题目、选项、正确答案和详细解析。

3.3.3. 个性化学习报告

评估智能体在生成学生个性化学习报告之前首先要对数据进行预处理，具体步骤如下：

1) 遍历所有知识点，将每个知识点的当前掌握度与班级平均掌握度进行对比，并给出“优秀”(>0.8)、“良好”(>0.6)或“薄弱”(≤0.6)的状态标签，形成知识点掌握情况清单。

2) 基于训练历史，统计每个知识点在训练过程中的平均增益值，将平均增益低于阈值(0.02)的知识点识别为“进步缓慢”项，形成错误分析文本。

上述预处理方法将数值化数据转化为自然语言段落，将学生智能体的多维状态以及最近 15 步的学习过程摘要(知识点、学习动作、掌握度变化)转换为结构化的文本描述作为提示词中的具体上下文，使大模型能够基于真实学情做出精准诊断。

3.3.4. 系统前端与演示

为直观展示评估智能体生成的个性化学习报告及相关数据，系统基于 Streamlit 构建了双区域响应式界面。用户选定待分析的学生后，主界面以三列卡片式指标展示各知识点的初始掌握度，侧边栏则解锁

“训练配置”模块，为后续训练做好准备。具体效果见图 4。

Figure 4. System front-end page—initial mastery status
图 4. 系统前端页面——初始掌握状态

完成指定训练轮次与每轮最大步数的训练后，主界面呈现训练前后掌握度对比数据表及柱状对比可视化图表，并汇总平均初始掌握度、平均最终掌握度与总体提升值，从而量化训练效果。具体效果如图 5 所示。

	知识点	初始掌握度	训练后掌握度	提升	状态
0	机器学习概论	0.993	0.996	+0.003	➡持平
1	线性回归	0.917	0.941	+0.025	➡持平
2	特征工程	0.833	0.993	+0.159	✅提升
3	朴素贝叶斯	0.900	1.000	+0.100	✅提升
4	模型评估	0.769	0.973	+0.203	✅提升
5	逻辑回归	1.000	0.897	-0.103	❌下降
6	决策树	1.000	0.929	-0.071	❌下降
7	聚类	0.875	0.979	+0.104	✅提升
8	集成学习	0.895	1.000	+0.105	✅提升
9	SVM	0.900	0.983	+0.083	✅提升

Figure 5. System front-end page—comparison of status before and after training
图 5. 系统前端页面——训练前后状态对比


机器学习课程个性化学习报告

学生姓名：
学生ID：
报告日期： 2026年1月

1. 学习效果总体评估

同学，你好！

首先，祝贺你在本学期的机器学习课程学习中取得了非常出色的整体成绩。你的**平均掌握度达到0.959**，这是一个相当高的水平，表明你对课程核心内容的理解非常扎实。尤其值得肯定的是，在**聚类、模型评估、集成学习**等多个重要且有一定难度的知识点上，你的掌握度（1.00，1.00，0.99）显著高于班级平均水平，这充分展现了你的**学习能力（0.97）**和**学习动机（1.40）**的优势。你能够通过挑战题等高阶任务来巩固和深化理解，这是一种非常高效的学习策略。

然而，在肯定成绩的同时，我们也需要客观地审视数据中反映出的潜在问题，这对于你未来的持续提升至关重要。

主要不足之处与分析：

1. **“高原现象”与学习效率问题**：从你的学习过程摘要来看，存在一个明显的现象：在多个知识点（如聚类、模型评估、决策树等）的学习后期，你的掌握度提升曲线变得非常平缓（例如，聚类连续7次挑战题掌握度无变化，模型评估后期每次仅提升约0.009）。这表明你可能遇到了“学习高原期”——即通过当前的学习方式（反复做同类题目）已经难以获得新的认知突破。你的高动机驱使你投入了大量练习，但**学习方法可能需要从“量的重复”转向“质的深化”**。
2. **相对薄弱环节**：虽然所有知识点掌握度均在“优秀”等级，但**“逻辑回归”（0.88）** 是你的相对短板，其掌握度低于班级平均（0.98）整整0.1个点。这需要引起特别关注。此外，

Figure 6. System front-end page—personalized student report
图 6. 系统前端页面——学生个性化报告

系统进一步生成个性化学习报告，涵盖学习效果总体评估、薄弱知识点深度分析、个性化学习建议、推荐学习策略、面向教师的教学策略建议以及针对性练习题六个核心模块，实现了从数值评价到可解释性反馈的转化，为学生和教师提供精准、深度的个性化支持。具体效果如图 6、图 7 所示。

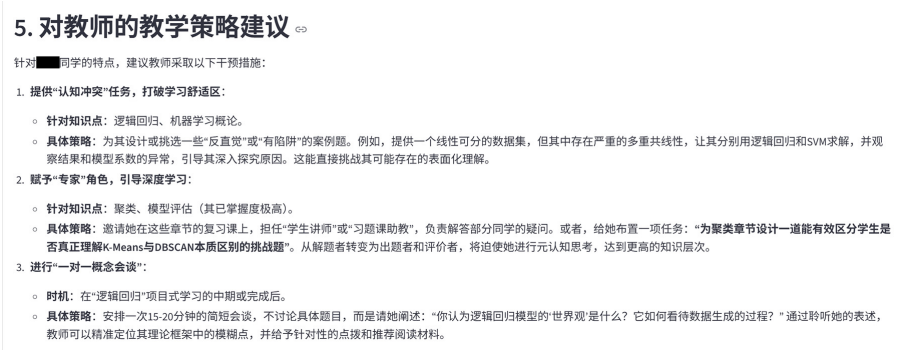

Figure 7. System front-end page—teacher strategy suggestions
图 7. 系统前端页面——教师策略意见

4. 实验与结果分析

4.1. 数据集描述

本系统所用数据集来源于某高校信息工程专业《机器学习》课程的真实教学答题记录，数据集共包含 6670 条答题记录，涵盖 51 名学生对 10 个核心课程单元的作答情况，数据维度包含学生 ID、姓名、问题 ID、问题内容、答案、分数及知识点类别 7 个字段。

数据内容聚焦机器学习领域，知识点类别覆盖“回归分析”、“强化学习”等核心模块，每个模块均包含选择题、判断题及主观问答题等多种题型。答题分数取值范围为 0~20 分，能够全面反映学生对各知识点的掌握程度。

4.2. 实验指标评估

4.2.1. 有效性分析

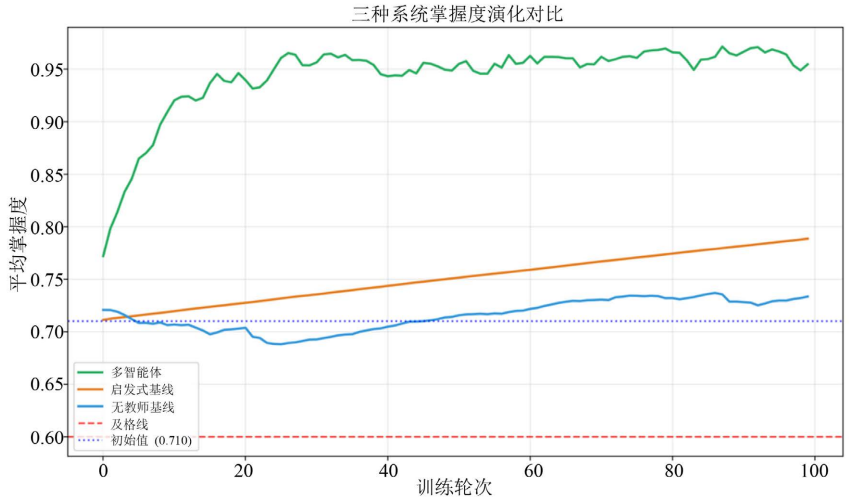

Figure 8. Comparison of mastery evolution in baseline system
图 8. 基线系统掌握度演化对比

为验证本文提出的多智能体协同强化学习框架的有效性，我们设计了两种基线系统进行对比实验：1) 启发式基线：教师智能体仅采用固定启发式规则，无学习能力。2) 无教师基线：移除教师智能体，学生智能体独立学习。三种系统均在相同的学生(初始平均掌握度 0.71)上训练 100 轮次。

启发式基线的最终掌握度仅升至 0.789，这表明缺乏自适应学习的固定教学策略难以充分激发学生潜力，验证了强化学习在教师策略优化中的必要性。而无教师基线进步也十分有限。学生独立学习最终掌握度仅达到 0.733。该结果进一步凸显了教师智能体在协同框架中的关键作用。

三种系统在训练过程中的平均掌握度变化曲线如图 8 所示。可见，多智能体系统在约 17 轮后快速收敛至高掌握度，而两种基线系统的掌握度始终处于低位或平缓状态，充分体现了多智能体系统在加速学习收敛和提升最终掌握度方面的有效性。

4.2.2. 创新点的实际贡献分析

为验证本文提出的学生模型中动机与疲劳度两个心理状态变量的实际贡献，设计消融实验。在保持教师智能体及其他参数不变的前提下，分别构建三种消融变体：1) 无动机：移除动机变量，动机因子恒为 1；2) 无疲劳度：移除疲劳度变量，疲劳度始终为 0；3) 无动机 + 无疲劳：同时移除两者。各变体与完整系统在同一学生(初始平均掌握度 0.71)上训练 100 轮次，实验结果如表 3 所示。

Table 3. Comparison of ablation experiment results
表 3. 消融实验结果对比

变体	最终平均掌握度	提升幅度	稳定性(σ)
完整系统	0.953	+0.243	0.0055
无动机	0.857	+0.147	0.0057
无疲劳度	0.838	+0.128	0.0052
无动机 + 无疲劳	0.753	+0.043	0.0105

消融实验结果表明，移除动机变量后最终掌握度降至 0.857，验证了动机对学生积极投入学习的激励作用；移除疲劳度后最终掌握度降至 0.838，反映了疲劳度机制在避免过度学习、维持高效状态方面的关键作用。同时移除动机与疲劳度后，最终掌握度进一步下降至 0.753，且学习过程波动剧烈、稳定性显著降低，表明两者之间存在协同效应：动机提供内在驱动力，疲劳度调节学习节奏，二者共同作用方能充分发挥教师智能体的引导能力。

5. 总结与展望

本文针对传统智能教学系统动态适应性不足、忽视学生认知情感状态及师生交互薄弱等问题，设计并实现了一种基于多智能体协同的动态教学优化系统。系统构建了“学生－教师－评估”三位一体的智能体框架，学生智能体基于 Q-Learning 算法，融合固有学习属性、动态学习状态与历史行为记录构建多维状态空间，自主选择最优学习动作；教师智能体通过离散化学生状态、融合启发式规则与强化学习，动态调整教学节奏、难度与策略；评估智能体以 DeepSeek 大语言模型为核心，生成结构化个性化学习报告，形成“教学－学习－评估”闭环。实验结果表明，该系统能有效提升学生知识点掌握度，且策略收敛稳定，教师决策符合真实教学规律，实现了规模化与个性化的教学统一。

未来研究将从三方面进一步优化：一是拓展状态空间维度，引入学生学习风格、知识遗忘曲线等因素，提升状态表征的精准度；二是优化多智能体协同机制，引入博弈论思想设计更灵活的交互策略，增强智能体间的动态适配能力；三是扩大数据集规模与场景覆盖，将系统应用于不同学段、学科的教学场

景，验证其泛化性，同时探索与知识图谱、自适应学习路径规划技术的融合，推动智能教学系统向更深度的个性化与智能化发展。

参考文献

[1] 郑娅峰, 黄璟玥, 包昊罡. 教育智能体赋能科学教育: 概念特征、应用价值与实施策略[J]. 远程教育杂志, 2025, 43(3): 24-32.

[2] 国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. 2017-07-20. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm, 2025-04-25.

[3] 教育部. 教育部等六部门印发意见部署教育新型基础设施建设[EB/OL]. 2021-07-21. http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202107/t20210721_545968.html, 2026-03-17.

[4] 邱燕楠, 李政涛. 挑战·融合·变革: “ChatGPT 与未来教育”会议综述[J]. 现代远程教育研究, 2023, 35(3): 3-12+21.

[5] 于济凡, 李睿淼, 李曼丽, 等. 多智能体协同交互的高临场感在线学习环境构建[J]. 现代教育技术, 2024, 34(12): 17-26.

[6] 吴永和, 姜元昊, 陈圆圆, 等. 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[J]. 开放教育研究, 2024, 30(5): 63-75.

[7] 郑娅峰, 赵亚宁, 黄璟玥, 等. 教育智能体: 研究现状和发展趋势[J]. 现代远程教育研究, 2025, 37(4): 3-13+59.

[8] 梁竹梅, 李鲍, 赵冬梅. 以 AI 智能体重构学习过程——教学智能体创建案例分析与思考[J]. 中国大学教学, 2025(9): 80-86.

[9] 赵德京, 马洪聪, 廖登宇, 等. 一种基于动作采样的 Q 学习算法[J]. 控制工程, 2024, 31(1): 70-79.

[10] 李明阳, 许可儿, 宋志强, 等. 多智能体强化学习算法研究综述[J]. 计算机科学与探索, 2024, 18(8): 1979-1997.

[11] 郭锐, 吴敏, 彭军, 等. 一种新的多智能体 Q 学习算法[J]. 自动化学报, 2007(4): 367-372.

[12] 林鸿生, 刘尚富, 赵磊. “AI 大模型 + 教师”人机协同教学策略研究[J]. 中国教育技术装备, 2026(3): 16-18+23.

[13] 仵政源, 赵诗奎, 解瑞建, 等. 融合贪心策略的遗传-禁忌搜索算法求解分布式装配作业车间调度问题[J/OL]. 中国机械工程, 1-16. https://link.cnki.net/urlid/42.1294.TH.20260121.1041.006, 2026-03-22.

[14] 蒋忠元, 陶梅悦, 赵晓庆, 等. 基于启发式规则的流式在线日志解析方法[J]. 通信学报, 2024, 45(4): 95-113.

[15] 王旭, 朱其新, 朱永红, 等. 改进 Q 学习算法的移动机器人路径规划[J]. 计算机仿真, 2025, 42(4): 371-377.