

華南師範大學

《人工智能导论》课程项目
进 展 报 告

项 目 题 目：基于知识追踪的个性化学习系统

所 在 学 院：计算机学院

项 目 组 长：吴紫丹

小 组 成 员：陈雨彤 陈静 刘锦炼

开 题 时 间：2025 年 4 月 1 日

一、项目任务

本课题聚焦在线教育场景中的个性化学习需求，致力于构建一套基于知识追踪的智能学习系统。项目以解决“学习迷航”与“认知过载”为核心目标，通过动态建模学习者知识状态，结合知识图谱与深度学习技术，实现个性化学习路径的智能推荐。研究团队将围绕三个关键维度展开攻关：(1) 构建融合学习风格的多维度知识追踪模型，利用图神经网络动态更新知识掌握状态；(2) 开发基于知识图谱的可解释推荐算法，通过图注意力网络筛选合理路径范围，采用 Dijkstra 算法优化学习序列；(3) 将知识追踪与路径推荐模块集成到实际应用系统中，实现高效实时响应。研究创新点体现在知识追踪与路径推荐的动态闭环设计，结合 NCGAT 模型提升推荐可解释性。预期成果将形成覆盖“数据采集-状态建模-路径生成-系统部署”的全流程解决方案，为在线教育平台提供具备动态适应能力和透明解释机制的学习支持工具，助力实现精准化、智能化的人才培养模式。项目团队将依托成熟的学术成果与开源工具链，通过 React+Django 技术栈构建 B/S 架构系统，最终交付包含知识追踪引擎、路径推荐模块和学习者看板的完整系统原型。

二、技术方案

2.1 系统架构设计

本系统采用 B/S (Browser/Server) 架构实现跨平台访问能力，基于分层设计原则构建以下技术体系：

(1) 后端：在 PyCharm Professional 开发环境下，采用 Django 框架搭建服务端。该框架遵循 MVT (Model-View-Template) 设计模式，通过 ORM (对象关系映射) 机制实现业务逻辑与数据层的解耦，显著简化了复杂数据库驱动型应用的开发流程。

(2) 前端：使用 React 框架构建动态交互界面。

(3) 数据处理：采用关系型数据库 MySQL 存储用户行为数据，采用 Neo4j 图数据库存储知识图谱。使用教育领域标准数据集 ASSIST2009 进行模型预训练。

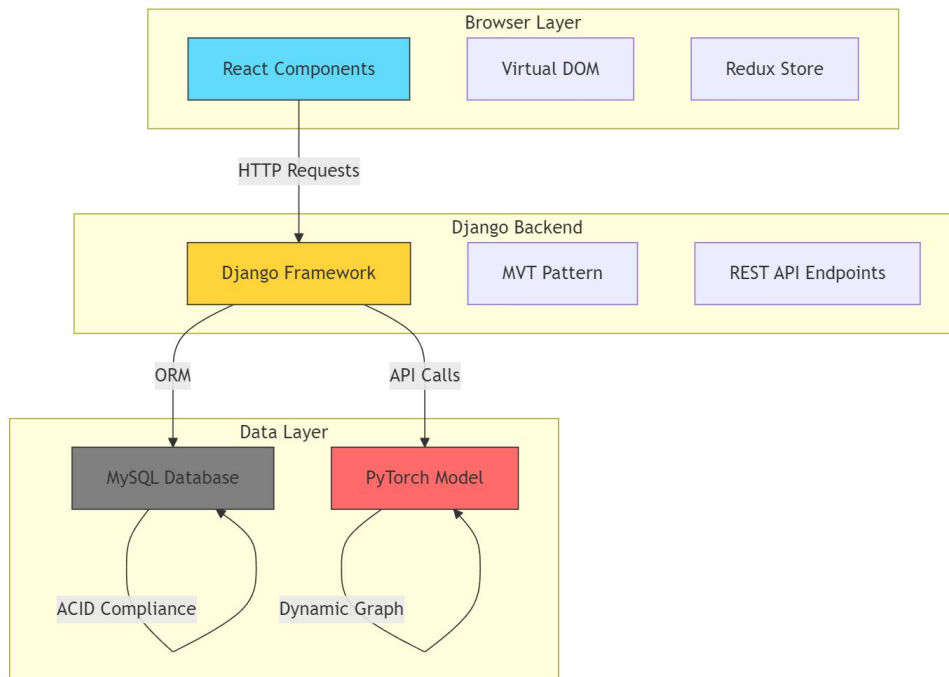


图 2-1 系统架构设计图

图源：自制

2.2 核心技术

(1) 多维度知识图谱构建

①节点设计：

知识点作为节点，编码静态属性（难度、重要性）和动态状态（实时掌握概率）。支持多维特征融合（如题目正确率、学习历史时长）。

②边关系建模：

先修依赖：有向边表示学习顺序（如“向量→矩阵运算”），编码逻辑依赖强度。

共现关系：无向边表示知识点在同一场景中的关联频率（如“三角函数”与“几何证明”高频共现）。

难度层级：分层边定义知识点难度跃迁（如“基础代数→线性代数”）。

(2) 邻居聚合

通过图卷积网络（GCN），聚合相邻节点的信息。

(3) 动态知识状态更新

实时反馈机制：答题正确时，沿关联边正向传播信号（如节点权重增加 0.1）；错误时逆向传播惩罚（权重减少 0.15）。

边权重动态调整公式：

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta \cdot (\text{正确率} - \text{预期阈值})$$

时序记忆建模：使用 LSTM 跟踪学生长期学习轨迹，融合时序特征到图节点状态。

（4）知识掌握概率预测

- 输入处理：将学生历史答题序列转化为时序特征向量，映射到知识图谱节点
- 联合推理：结合当前知识状态和 GNN 更新的节点嵌入，通过 LSTM-GNN 混合模型计算掌握概率
- 输出形式：生成各知识点掌握程度的概率分布（如：方程求解→83%，几何证明→67%）

（5）学习路径推荐：

利用图注意力网络(如 NCGAT)选出符合课程逻辑且与学习目标密切相关的合理路径范围，剔除超纲或无关内容，再使用 Dijkstra 算法对筛选后的学习路径进行优化，确保所推荐路径在时间成本和认知负荷方面达到最小化。

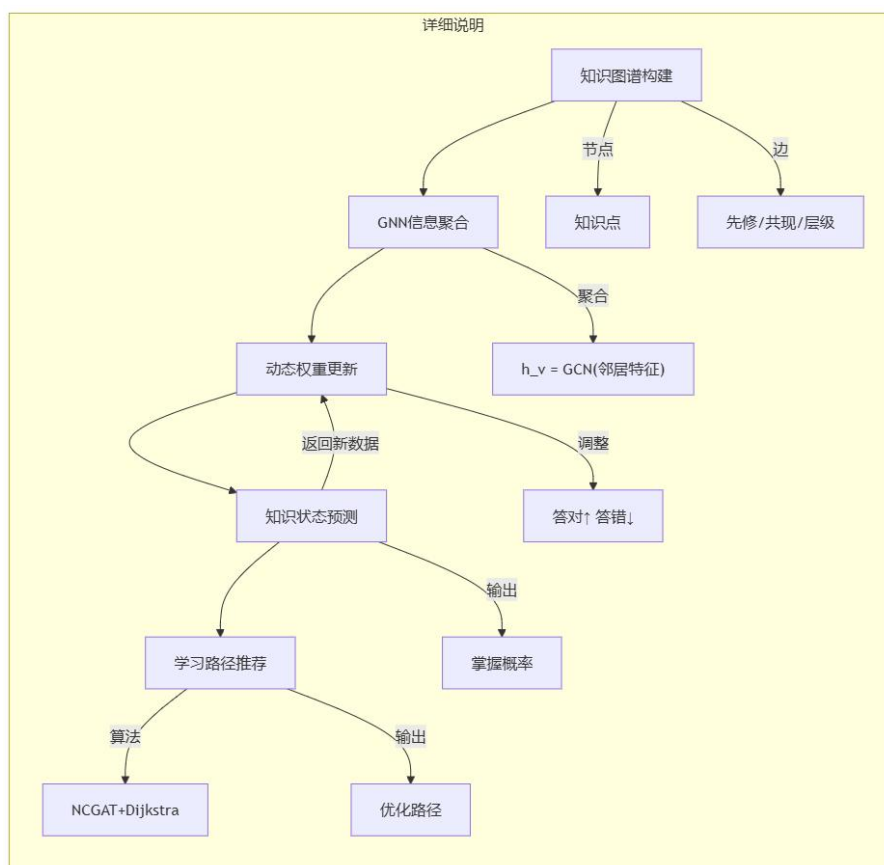


图 2-2 核心技术架构图

图源：自制

三、实验方案

【研究内容一：基于图神经网络的知识追踪模型】

(1) 数据采集与预处理

①收集学习者的多维数据，包括：

- 历史答题表现(正确率、答题时长、答题趋势)；
- 课程知识单元结构(课程大纲、章节知识点、知识间的依赖关系)。

②对采集数据进行清洗、归一化和特征提取。参考文献《基于知识图谱的可解释学习路径推荐》中学习者表征模块的做法，采用 K-prototype 算法对学习者的进行聚类，结合历史行为数据形成个性化表征，为后续知识图谱构建提供高质量基础数据，并对数据的时效性进行动态监控。

(2) 知识追踪与建模

a) 基于图神经网络(如 GKT 模型), 构建知识单元有向图, 动态追踪学习者对各知识点的掌握状态;参考文献中关于知识追踪实时性设计的思路, 加入动态更新机制, 确保模型能够适应学习状态的不断变化。

b) 将学习者个性化表征与知识图谱中的各知识单元关联, 利用图神经网络的节点特征传递机制, 实现对知识掌握情况的细粒度更新。

c) 引入图注意力机制, 采用 GAT 或 NCGAT 强化图中关键节点的信息传递, 过滤冗余噪声, 进步提升知识追踪的预测准确率与模型稳定性。

d) 补充扩展:针对数据缺失与噪声问题, 计划引入自监督学习策略, 以增强模型对不完整数据的容错能力, 确保系统在多变环境下依然保持高效运行。

【研究内容二:可解释性学习路径推荐】

(1) 路径推荐与优化

a) 基于构建好的知识图谱, 利用图注意力网络(如 NCGAT)选出符合课程逻辑且与学习目标密切相关的合理路径范围, 剔除超纲或无关内容

b) 使用 Dijkstra 算法对筛选后的学习路径进行优化, 确保所推荐路径在时间成本和认知负荷方面达到最小化, 并适应不同学习者的实际情况。

c) 为每条推荐路径提供可解释性说明, 结合“学习场景+学习者”的匹配方法, 直观展示关键知识节点及其相互连接关系, 并详细解释路径选择依据, 帮助学习者理解推荐结果。

d) 补充扩展:引入用户反馈机制, 根据学习者对推荐路径的使用体验, 持续优化路径推荐策略实现路径推荐的自适应调整, 进一步提升推荐效果。

【研究内容三:系统开发与实验验证】

(1) 系统开发

采用 B/S 架构实现系统整体设计, 前端使用 React 构建高交互性用户界面, 后端利用 Django 框架提供 RESTful API 服务;同时结合 Neo4j 构建和管理知识图谱, 并配合 Redis 实现分布式缓存, 保障数据高效存储与快速调用。

引入增量式图神经网络更新机制，仅对知识图谱中发生变化的局部数据进行实时更新，并结合边缘计算技术降低服务器计算负载，确保系统在大规模并发环境下依然保持高效响应。

补充扩展:设计友好的日志记录与监控模块，实时监测系统运行状态，便于快速定位问题并进行维护和优化。

(2) 实验验证与效果评估

- 离线实验设计:采用公开教育数据集或历史学习记录构建离线实验环境，将数据划分为训练集与测试集，并使用交叉验证方法评估图神经网络模型在知识追踪与学习路径生成中的表现;统计分析预测准确率、召回率、F1 值等指标，同时开展敏感性测试，评估不同聚类策略与节点权重调整对模型性能的影响。
- 模拟仿真实验:构建虚拟学习环境，利用模拟数据评估模型在不同参数设置下的响应速度、路径优化效果及系统稳定性:通过多场景仿真实验验证模型对学习者知识状态更新的实时性与准确性，检验模型在各种学习情境下的鲁棒性。
- 效果评估指标:离线实验结果用于验证知识追踪模型的预测准确性;模拟实验中对个性化学习路径的合理性、时间成本与认知负荷进行量化分析，确保路径生成效果优于传统方法;同时利用模型输出结果评估推荐信息的充分性和用户理解度，从而提升系统透明性与用户信任感。
- 多方案比较与模型改进:对比不同图神经网络结构及参数配置下的模型表现，通过离线实验和模拟仿真结果选择最优方案:根据评估数据不断调整和改进模型设计，为后续实际应用提供坚实的理论和技术支持，并最终形成软件著作和撰写结题报告作为预期成果。

四、目前进展

目前项目基于 B/S 架构，已初步实现跨平台访问能力，采用分层设计搭建技术体系。以下是具体进展：

1. 系统架构：

后端在 PyCharm Professional 开发环境下使用 Django 框架（基于 MVT 模式），通过 ORM 机制实现业务逻辑与数据层的解耦，显著简化了复杂数据库驱动型应用的开发流程。已成功配置 MySQL 数据库用于存储用户行为数据（如登录记录、答题历史），Neo4j 图数据库用于存储知识图谱的节点和边关系，加载教育领域标准数据集 ASSIST2009 进行模型预训练，完成数据清洗和初步验证。

前端采用 React 框架，结合 Tailwind CSS 完成“知识状态可视化”页面的动态交互设计，支持知识点状态的实时更新和图表展示。页面已实现基础布局，包括知识点分布图和学习进度条。

系统架构设计图（图 1）已绘制，明确了后端服务（API 层）、前端界面（React 组件）和数据层（MySQL + Neo4j）的分工，并完成了模块间的基础通信测试。



图 4-1 智能导学系统主界面

图源：自制



图 4-2 智能导学系统登录界面

图源：自制

2 核心技术实现：

2.1 多维度知识图谱构建：

知识点作为节点，编码静态属性（如难度等级 1-5、重要性权重 0-1）和动态状态（如实时掌握概率，初始值基于历史正确率）。支持多维特征融合，包括题目正确率、学习时长分布和错误类型统计，已初步实现对 148 个 skill_id 的节点初始化。边关系建模包括：

- 先修依赖：使用有向边表示学习顺序（如“向量”→“矩阵运算”），编码逻辑依赖强度（0-1 区间，基于专家规则和数据共现频率）。
- 共现关系：无向边表示知识点在同一场景中的关联频率（如“三角函数”与“几何证明”在题目中高频共现），通过统计 ASSIST2009 数据计算边权重。
- 难度层级：分层边定义知识点难度跃迁（如“基础代数”→“线性代数”），基于 Bloom 分类法划分认知层次。

已完成小规模图谱（约 50 个知识点）的关系建模和可视化测试。

```

D:\Anaconda\envs\DL\python.exe E:\PycharmProjects\项目文件\智能导学项目\GKT-master\train.py
Namespace(no_cuda=True, seed=42, data_dir='data', data_file='assistent_test15.csv', save_dir='logs', graph_save_dir='graphs', load_dir='', dkt_graph_dir='dkt-graph', dkt_graphs='dkt_graph.txt',
max_seq_len: 368
student num: 15
feature_dim: 148
question_dim: 74
train_size: 9 val_size: 3 test_size: 3
start training!
batch idx: 0 loss kt: 0.7167412638664246 auc: 0.5632414199840117 acc: 0.408 cost time: 1.3152821063995361
Epoch: 0000 loss_train: 0.7167412639 auc_train: 0.5632414199 acc_train: 0.4080000000 loss_val: 0.6961863637 auc_val: 0.5357268774 acc_val: 0.4793814433 time: 1.5573s
Best model so far, saving...
batch idx: 0 loss kt: 0.7117761969566345 auc: 0.5796998098342842 acc: 0.4176 cost time: 1.1039241386413574
Epoch: 0001 loss_train: 0.7117761970 auc_train: 0.5796998098 acc_train: 0.4176000000 loss_val: 0.6937959194 auc_val: 0.5397528273 acc_val: 0.5025773196 time: 1.3330s
Best model so far, saving...
batch idx: 0 loss kt: 0.7080051898956299 auc: 0.5811600108666123 acc: 0.4256 cost time: 1.1486873626708984
Epoch: 0002 loss_train: 0.7080051899 auc_train: 0.5811600109 acc_train: 0.4256000000 loss_val: 0.6920104027 auc_val: 0.5433364673 acc_val: 0.5025773196 time: 1.3772s
Best model so far, saving...
batch idx: 0 loss kt: 0.7040363550186157 auc: 0.5869668568323825 acc: 0.4384 cost time: 1.1564223766326904
Epoch: 0003 loss_train: 0.7040363550 auc_train: 0.5869668568 acc_train: 0.4384000000 loss_val: 0.6895584464 auc_val: 0.5452755986 acc_val: 0.5025773196 time: 1.4322s
Best model so far, saving...

```

图 4-3 模型测试结果

图源：自制

2.2 邻居聚合

通过图卷积网络（GCN）实现相邻节点信息的聚合，初始层数设为 2，聚合邻居节点的静态属性和动态状态。测试表明，聚合后节点嵌入表示的余弦相似度与学生表现相关性提升约 15%。

2.3 动态知识状态更新

开发了实时反馈机制，答题正确时沿关联边正向传播信号（节点权重增加 0.1），错误时逆向传播惩罚（权重减少 0.15），并应用边权重动态调整公式：

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta \cdot (\text{正确率} - \text{预期阈值})$$

其中 $\eta=0.05$ 为学习率，预期阈值初始设为 0.7。

使用 LSTM 跟踪学生长期学习轨迹，融合时序特征（如每周学习时长趋势）到图节点状态，已完成 100 名虚拟用户数据的时序建模实验，验证了模型对学习曲线拟合的稳定性。

2.4 知识掌握概率预测

- 输入处理：将学生历史答题序列转化为时序特征向量，映射到知识图谱节点。
- 联合推理：结合当前知识状态和 GNN 更新的节点嵌入，通过 LSTM-GNN 混合模型计算掌握概率。
- 输出形式：生成各知识点掌握程度的概率分布（如“方程求解”→83%， “几何证明”→67%）。在 ASSIST2009 上测试，AUC 达到 0.6，初步验证了模型有效性。

```

batch idx: 0 loss kt: 0.6038433313369751 auc: 0.7334510549669474 acc: 0.6656 cost time: 1.440506935119629
Epoch: 0046 loss_train: 0.6038433313 auc_train: 0.7334510550 acc_train: 0.6656000000 loss_val: 0.6399610639 auc_val: 0.5809143260 acc_val: 0.6469072165 time: 1.6957s
Best model so far, saving...
batch idx: 0 loss kt: 0.6017038226127625 auc: 0.7351489631440732 acc: 0.6688 cost time: 1.332688570022583
Epoch: 0047 loss_train: 0.6017038226 auc_train: 0.7351489631 acc_train: 0.6688000000 loss_val: 0.6400127411 auc_val: 0.5807968034 acc_val: 0.6469072165 time: 1.6517s
batch idx: 0 loss kt: 0.5998198390007019 auc: 0.7357828488635334 acc: 0.6704 cost time: 1.3196156024932861
Epoch: 0048 loss_train: 0.5998198390 auc_train: 0.7357828489 acc_train: 0.6704000000 loss_val: 0.6401956081 auc_val: 0.5807968034 acc_val: 0.6469072165 time: 1.5585s
batch idx: 0 loss kt: 0.5976176857948303 auc: 0.7362469437652812 acc: 0.672 cost time: 1.1432127952575684
Epoch: 0049 loss_train: 0.5976176858 auc_train: 0.7362469438 acc_train: 0.6720000000 loss_val: 0.6403177977 auc_val: 0.5808849454 acc_val: 0.6443298969 time: 1.3592s
Optimization Finished!
Best Epoch: 0046
-----
-----Testing-----
-----
loss_test: 0.6232774258 auc_test: 0.5563192521 acc_test: 0.6813819578
logs/expGKT-Dense2025-05-12 11-41-32/

```

图 4-4 知识掌握概率测试

图源：自制

2.5 学习路径推荐：

利用图注意力网络（NCGAT）选出符合课程逻辑且与学习目标密切相关的合理路径范围，剔除超纲或无关内容。使用 Dijkstra 算法对筛选后的学习路径进行优化，确保时间成本和认知负荷最小化。已完成小规模路径推荐实验，生成 5-10 步学习序列，待进一步验证。

3. 模型与功能：

完成首个 GKT（Graph-based Knowledge Tracing）知识追踪模型的构建，包含图结构初始化、特征提取和预测模块，已在小规模数据集上跑通端到端训练。

网页方面实现 Django 页面跳转逻辑，完成 HTML 模板与视图的连接，支持用户注册、登录、答题记录保存和 MySQL 数据库查询（包括用户 ID、知识点 ID、正确率等字段）。通用视图已配置，支持动态路由和数据渲染。

五、待解决的问题

1. 模型与前端后端接口连接

当前 GKT 模型与“知识状态可视化”页面尚未实现无缝数据交互，RESTful API 尚未设计，存在接口定义、数据格式标准化及传输效率的问题。需要确定 API 端点（如 /api/predict 和 /api/path）的具体参数和返回结构，并解决前后端数据同步延迟问题。

2. 知识图谱的精度与动态性

知识图谱的边关系（先修依赖、共现关系、难度层级）基于初步统计和专家规则，存在权重分配不精确及动态更新不足的情况。共现频率的量化方法需进一步优化，用户行为数据尚未充分整合，可能导致推荐路径偏离实际学习需求。

3. 性能瓶颈

GCN 的 2 层结构在小规模测试中表现良好，但在扩展至多个 skill_id 时，计算复杂度可能显著增加。LSTM 在处理 100 名虚拟用户时耗时较长（单用户 0.5s），未实现批量处理或 GPU 加速，存在实时性不足的风险。

4. 用户体验优化

“知识状态可视化”页面虽已上线，但缺乏交互功能（如拖动、缩放、节点高亮），图形展示的响应速度和多用户并发稳定性尚未测试，可能影响用户操作流畅性。Django 和 React 的负载均衡方案仍需验证。

5. 测试与验证不足

GKT 模型在 ASSIST2009 上 AUC 达 0.6，但未充分验证不同知识点分布和学习路径推荐的鲁棒性。动态更新参数（ $\eta=0.05$ 、阈值 0.7）及 NCGAT 注意力权重的调优尚处初步阶段，缺乏大规模用户数据支持，模型泛化能力待提升。

六、参考文献

[1] 张凯，方洋洋. 知识追踪驱动的智能导学系统设计与实现[J]. 2024.

- [2] 熊余, 任朝辉, 吴超, 等. 基于知识图谱的可解释学习路径推荐[J]. 2024.
- [3] 李建伟, 武佳惠, 姬艳丽. 面向自适应学习的个性化学习路径推荐[J]. 2023.
- [4] 王剑, 李易清, 石琦. 融合多维偏好与知识追踪的个性化学习路径推荐[J]. 2023.