

基于大数据的个性化智慧学习环境 构建方法研究



于 戈
东北大学计算机学院

提纲



一 研究背景

3

□ 时代需要新一代教学体系

新型教育体系：“**网络化、数字化、个性化、终身化**”

新型教学模式：“**一生一空间、生生有特色**”

学习型社会：“**人人皆学、处处能学、时时可学**”

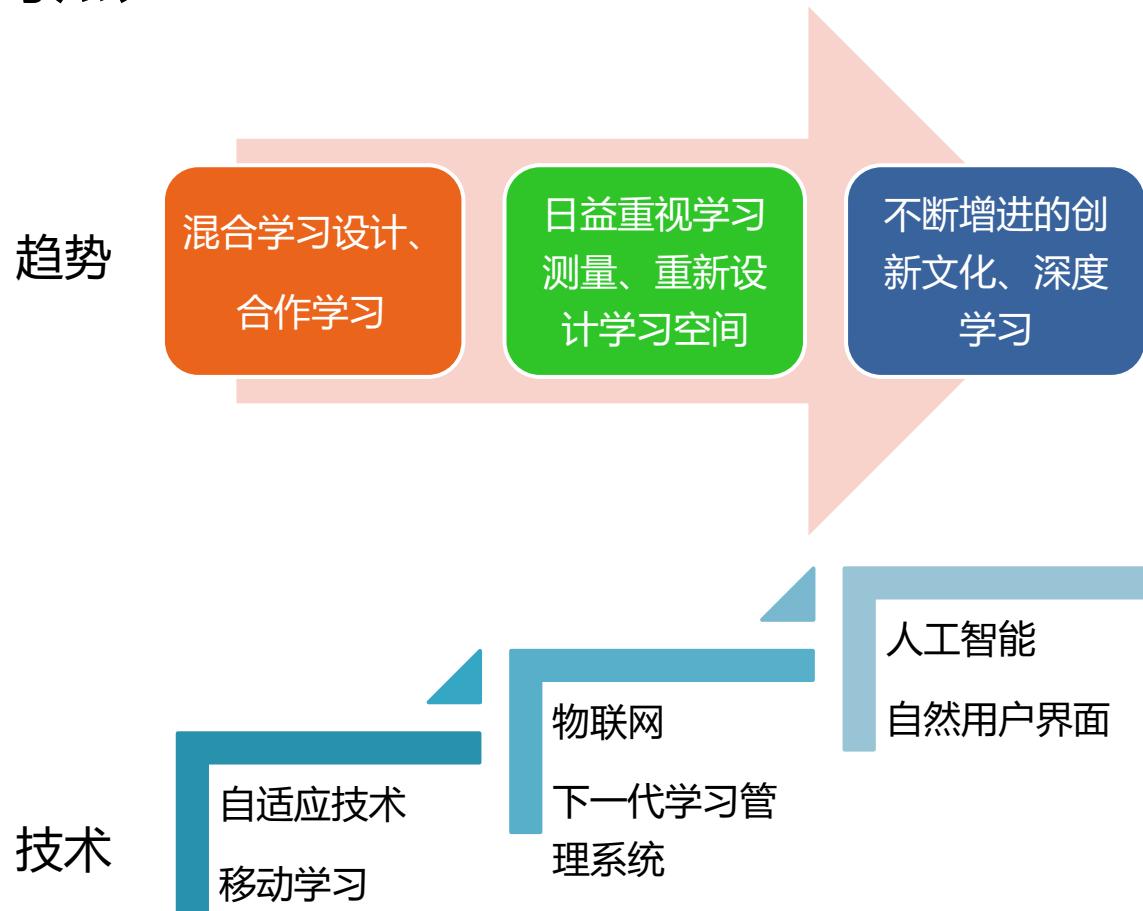
- 《教育部关于印发《教育信息化“十三五”规划》的通知》教技〔2016〕2号



一 研究背景

□ 新一代教学体系特点

- 学习需求的**个性化**
- 教学资源的**数字化**
- 知识传播的**网络化**
- 学习空间的**泛在化**
- 教学交互的**社交化**
- 教学支持的**智能化**



一 研究背景

5

- 2017-2020年教育技术发展回顾
 - 2020年地平线报告(教与学版)
 - Adaptive Learning Technologies
 - AI/Machine Learning Education Applications
 - Analytics for Student Success
 - Elevation of Instructional Design, Learning Engineering, and UX Design in Pedagogy
 - Open Educational Resources
 - XR (AR/VR/MR/Haptic) Technologies
 - 2021年地平线报告(教与学版)
 - Artificial Intelligence (AI)
 - Blended and Hybrid Course Models
 - Learning Analytics
 - Micro credentialing
 - Open Educational Resources (OER)
 - Quality Online Learning



- Artificial Intelligence: Technology Implications
- Next-Generation Digital Learning Environment (NGDLE)
- Analytics and Privacy Questions



- Widespread Adoption of Hybrid Learning Models
- Increased Use of Learning Technologies
- Online Faculty Development

一 研究背景

□ 教学环境定义

建构主义、联通主义、系统论对教学环境给出不同的定义

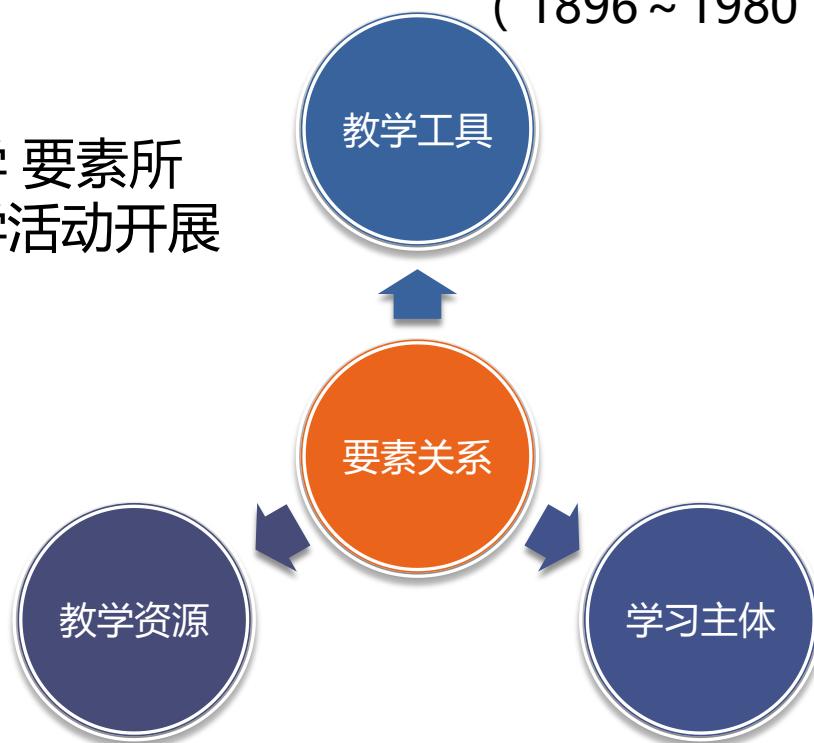


皮亚杰
(1896 ~ 1980)

教学环境：一个由紧密联系的多种教学 要素所组成，能够有效支撑教学活动开展的复杂系统

教学要素：工具、资源、学习主体

要素关系：知识结构、交互关系
共同体关系等



一 研究背景

7

□ 教学环境的变化及特点

以教师为主的
传统教学环境

基于传统信息技术的
数字化学习空间

非连续的
离散型学习空间

以学生为主的
自主学习环境

基于大数据和AI的
智慧学习空间

基于移动计算和物联
网的泛在学习空间

特点：协作式、情景化、个性化、智能化、泛在化、网络化

一 研究背景

□ 新一代教学环境的要求

□ 教学的泛在性

- 需求：**时时处处学习** 线上/线下、课堂/课外、室内/室外
- 方案：**构建一体化的教学空间**

□ 认知的复杂性

- 需求：**深度学习** 生生、师生、学生-教学资源
- 方案：**构建更为有效的教学交互**

□ 学习的差异化

- 需求：**因材施教** 兴趣、需求、智力、精力
- 方案：**提供个性化的教学供给**

一 研究背景

9

□ 新一代教学环境的挑战

缺乏一体化的智能教学空间

教学空间分散存在、独立管理、不同归属、数据多源异构

物理空间与虚拟空间不能有机融合

学习情境及状态数据缺少有效整合

缺乏有效的教学交互

个体-社群呈多元化组织结构、多元化交互过程

交互过程缺乏适应性

交互设计缺乏优化

缺乏个性化的教学供给

学习者的基础、目的、时间等学习多样性

学习资源无差别供给状态、个性化推荐精度低

缺少个体教学模式定制

一 研究背景

10

□ 国内外研究现状

研究内容

教学空间构建

教学交互

个性化导学

一体化教学环境系统

存在问题



缺乏支持教学空间一体化的融合与管理技术

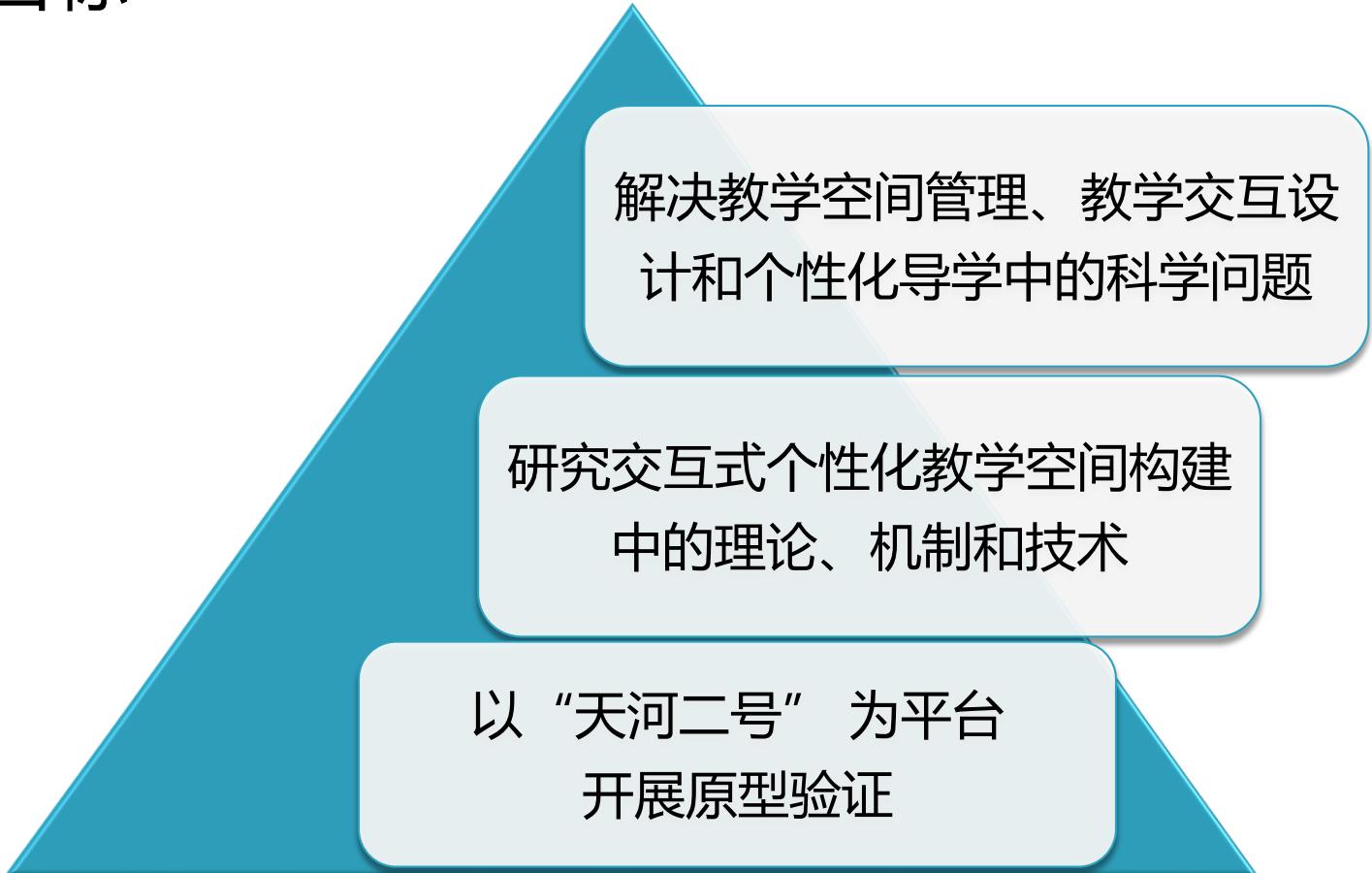
缺乏支持有效教学交互的设计方法和优化技术

缺乏满足差异性的个性化教学供给技术

与大数据、人工智能等技术的集成程度低

一 研究背景

□ 研究目标



解决教学空间管理、教学交互设计和个性化导学中的科学问题

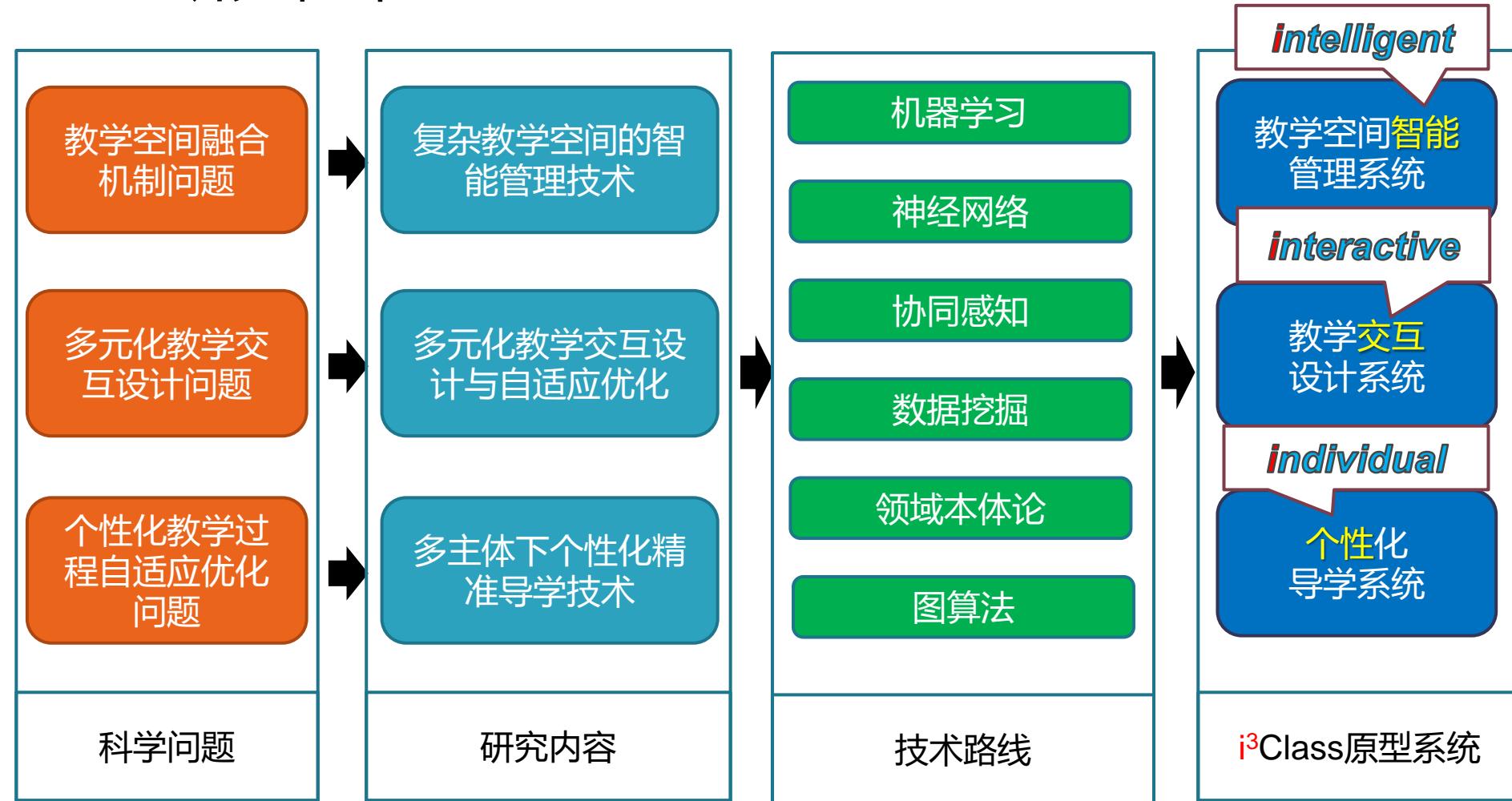
研究交互式个性化教学空间构建中的理论、机制和技术

以“天河二号”为平台
开展原型验证

二 研究内容

12

□ 研究框架



二 研究内容

13

□ 科学问题

问题1：多源异构教学空间融合机制问题

针对教学要素跨空间、多源、异构特点，解决：

- 一体化空间信息模型建立
- 教学过程数据融合
- 跨空间互操作



二 研究内容

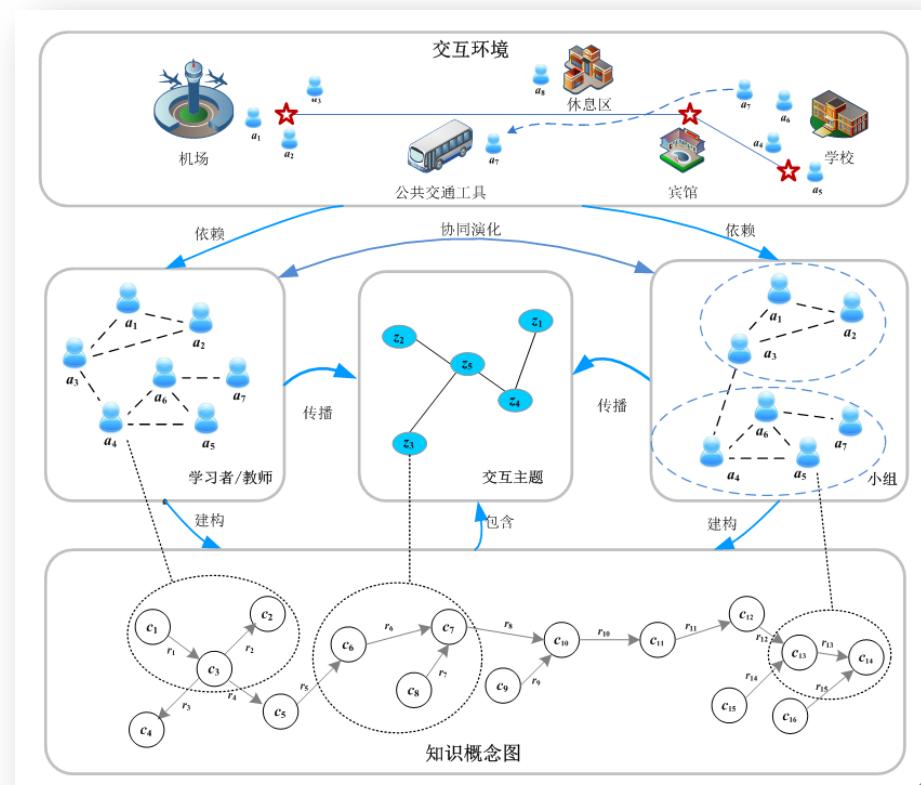
14

□ 科学问题

问题2：异构学习网络下的多元化教学交互设计问题

针对学习主体、学习资源和知识构成的异构网络，解决：

- 教学交互建模与质量评价
- 交互活动推荐
- 交互计划生成



二 研究内容

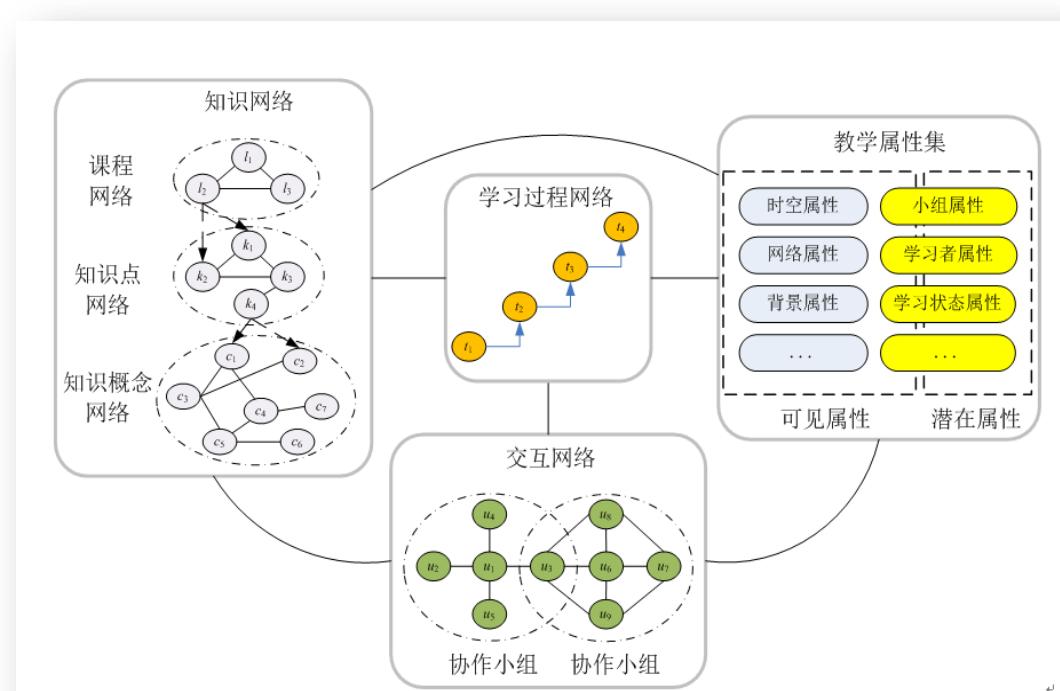
15

□ 科学问题

问题3：海量状态空间下的个性化教学过程自适应优化问题

针对学习状态空间庞大、教学评价不确定等复杂教学过程，解决：

- 个性化学习计划自适应优化
- 教学模式自适应优化



二 研究内容

□ 研究内容1：复杂教学空间的智能管理技术

重点

□ 多源异构教学空间融合机制

- 一体化空间信息模型与数据融合机制
- 基于事件处理的虚实空间互操作

□ 复杂环境下的智能情境感知技术

- 面向泛在学习空间的协同感知方法
- 面向协作学习的协作者和协作小组感知方法

□ 个性化教学空间构建技术

- 面向个体的个性化教学空间构建方法
- 面向群组的个性化教学空间构建方法

难点

跨空间的教学数据和事件的映射

多源行为感知、识别与融合

学习期望差异下的教学空间构建

二 研究内容

18

□ 研究内容2：多元化教学交互设计与自适应优化

重点

□ 教学交互模型及评价

- 教学交互建模
- 多元化交互特性分析
- 教学交互质量评价方法

□ 教学交互生成与自适应优化

- 教学交互活动推荐方法
- 教学交互计划生成方法
- 教学交互计划优化方法

难点

教学交互中知识结构、关系网络和互动拓扑特性的度量与分析

教学交互评价指标体系

综合交互主题、交互环境和交互主体的教学交互生成

二 研究内容

□ 研究内容3：多主体下个性化精准导学技术

重点

□ 教学内容资源多维推荐方法

- 教学内容多维推荐映射与参数学习
- 基于过程的推荐精炼方法

□ 大规模个性化学习计划自适应优化方法

- 基于马尔科夫决策过程的学习计划优化方法
- 基于教学效果反馈的学习计划自适应优化

□ 基于领域指导的个性化教学模式生成方法

- 基于教学领域理论的教学模式建模与教学模式挖掘方法
- 基于教学模式本体的个性化教学模式生成方法

难点

导学数据要素的
多维相似性度量

学习计划的计算复杂度和
转移代价确定

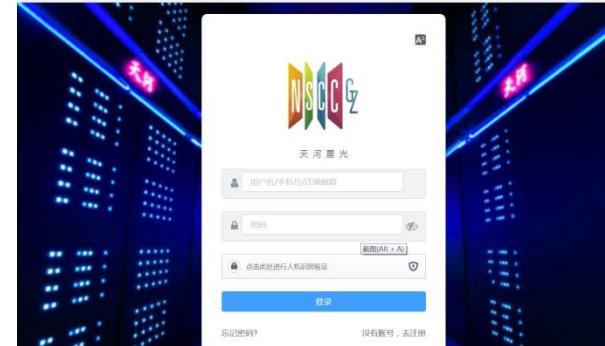
教学模式领域知识与
模式挖掘的结合

二 研究内容

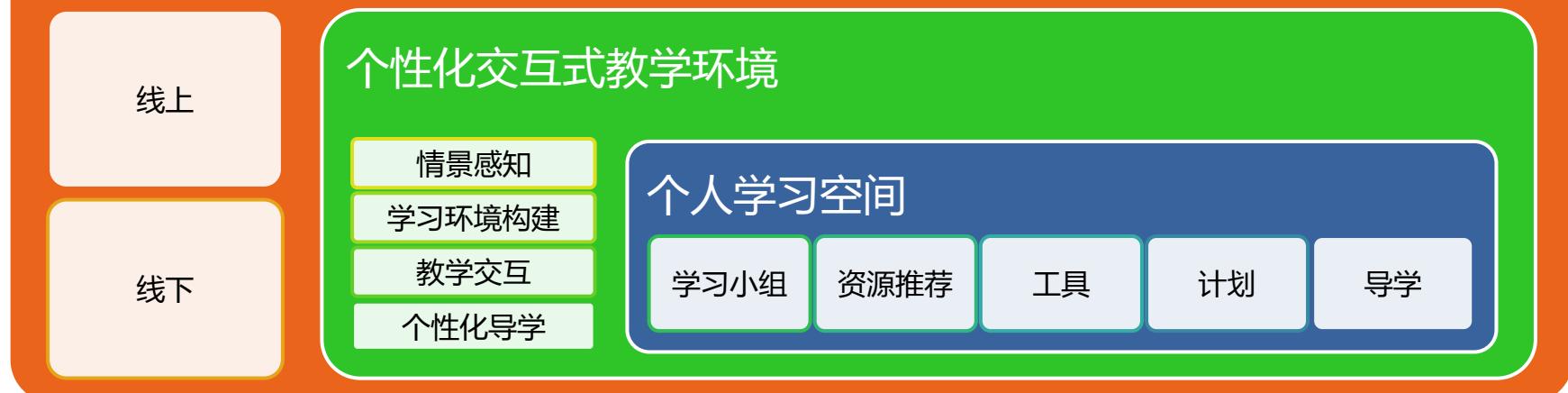
□ 研究内容4：交互式个性化教学环境原型系统

基于天河二号平台和现有在线学习系统

- 研制i³Class原型系统
- 面向校园和社会的各类学习者，建立个人学习空间，开展支持泛在学习的示范应用



学习型社会



二 研究内容

- 基于天河二号平台，研制原型系统i³Class
- 对现有的e-learning系统和MOOC系统集成，进行应用系统的开发与部署，形成i³Class示范性应用
- 基于课题组所在单位负责的辽宁省资源共享课、辽宁省跨校修读学习平台开展验证工作



二 研究内容

24

□ 特色与创新

问题

个性化交互式
教学环境构建理论

复杂教学空间
管理问题

有效的多元化
交互问题

多主体个性化
导学问题

创新

大数据驱动教学环境构建的
计算模型

基于一体化空间信息模型的
智能情境感知技术

基于异构交互网络模型的
教学交互设计方法

基于多维教学数据模型的
学习计划自适应优化算法

效果

支持新建构主义、联通主义等
先进教育理论

解决虚实分离、空间独立、信息
分散等问题，形成一体化空间

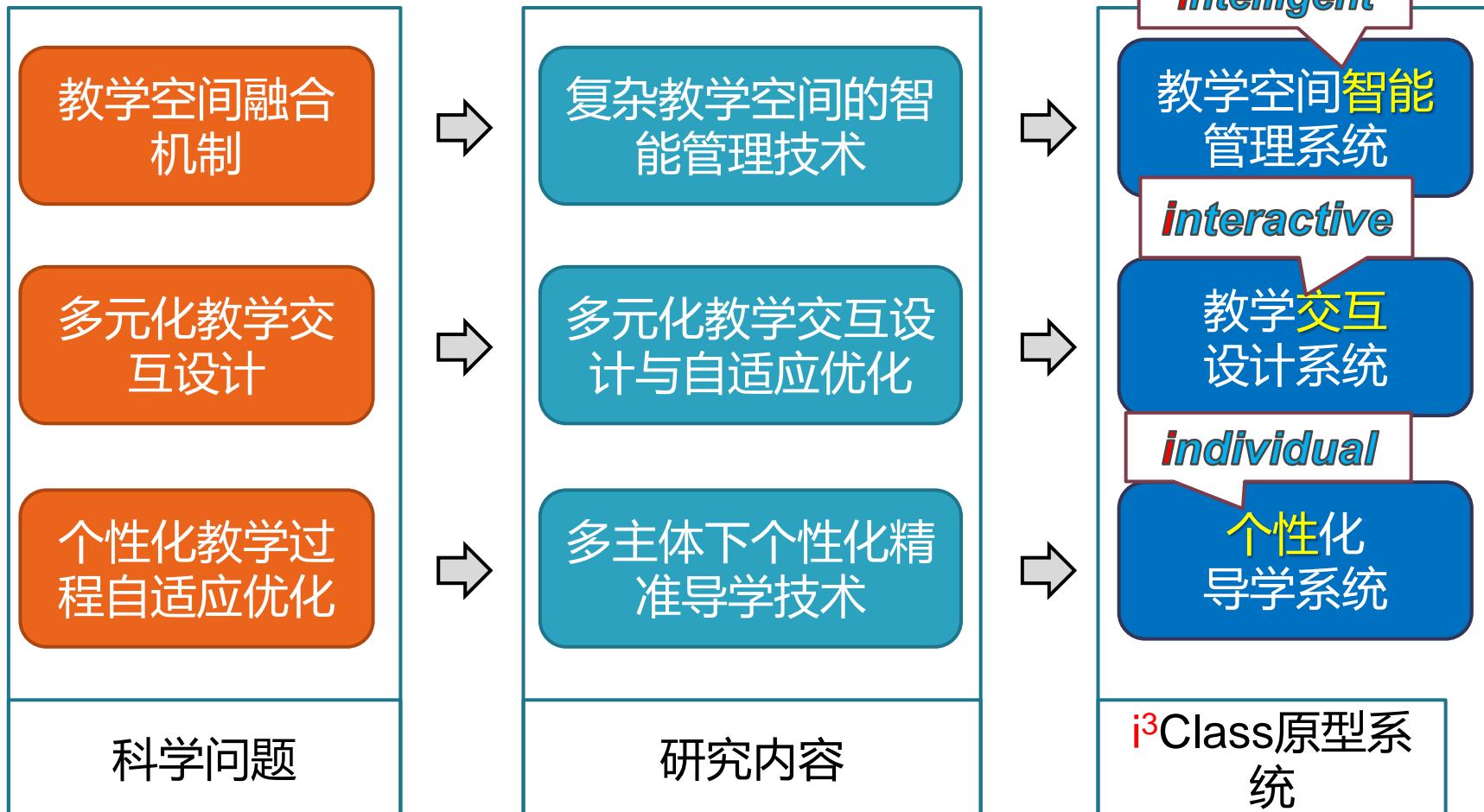
解决交互形式单一，满足按课程、
情景等不同场景的交互活动组织

解决推荐精度低、适应性差等问
题，形成导学资源精准定制

三 研究进展

25

□ 研究框架



研究进展1：复杂教学空间的智能管理技术

重点

□ 多源异构教学空间融合机制

- 基于事实检测的语义层次图推理方法
- 基于字符识别的试卷文本检测方法
- 异构关联数据一致性规则发现方法
- 基于AR学习环境的学习者行为采集与分析模型
- 基于多元图融合的异构信息网络嵌入方法

□ 复杂环境下的智能情境感知技术

- 基于LISTWISE监督的图像检索方法
- 基于金字塔哈希的大规模图像检索方法
- 基于顺序感知重加权策略的图像检索方法
- 基于集成学习的学习状态感知
- 教育场景下目标检测技术
- 基于关键表情区的面部表情识别方法研究

□ 个性化教学空间构建技术

- 基于事件的深层多智能体强化学习方法
- 个性化学习影响因素分析

成果

1. Wanjun Zhong, Duyu Tang, Zenan Xu, Ruize Wang, Nan Duan, Ming Zhou, Jiahai Wang, Jian Yin: Neural Deepfake Detection with Factual Structure of Text. EMNLP 2020.

2. Linteng Wang, Xiaoguang Li. Examination Paper Text Detection Based on Character Discriminator. ICDIP 2019.

3. 杜岳峰, 李晓光, 宋宝燕. 面向异构关联数据的一致性规则发现方法. 计算机研究与发展. 2019.

4. 李封, 张昱, 陈默 . 基于AR技术的学习者在线学习行为研究初探. 中国教育信息化. 2019.

5. 吴瑶; 申德荣*; 寇月; 聂铁铮; 于戈; 多元图融合的异构信息网嵌入, 计算机研究与发展, 2020.

1. Yang Y, Geng L, et al. Feature Pyramid Hashing. ICMR2019

2. Yang Y, Geng L, Lai H, et al. Deep Policy Hashing Network with Listwise Supervision. ICMR2019

3. Lai H, et al. Improving Deep Binary Embedding Networks by Order-aware Reweighting of Triplets. TCSV 2019

4. 邵馨莹. 基于集成学习的学习状态感知技术研究 . 硕士学位论文. 2019.

5. 范宇翔, 教育场景下目标检测技术研究 . 硕士学位论文. 2019.

6. 高猛, 基于关键表情区的面部表情识别方法研究. 硕士学位论文. 2019.

1. Zhenyu Shi, Runsheng Yu, Xinrun Wang, Rundong Wang, Youzhi Zhang, Hanjiang Lai, Bo An: Learning Expensive Coordination: An Event-Based Deep RL Approach. ICLR 2020.

2. Zhang T, Dong W, Shi H, et al. Design and implementation of teaching analysis system based on data mining. CCDC2019.

3. Li Feng, Zhang Yu, Chen Mo, Gao Kening. Which Factors Have the Greatest Impact on Student's Performance. Journal of Physics: Conference Series.2019

研究进展2：多元化教学交互设计与自适应优化

27

重点

□ 教学交互模型及评价

- 基于自定义预训练的多方向响应选择图推理网络
- 生成多跳推理问题以提高机器阅读理解能力
- 支持移动协作学习的异构交互模型构建
- 基于学习者行为的教学交互的度量与分析

成果

1. Yongkang Liu, Shi Feng, Daling Wang, Kaisong Song, Feiliang Ren, Yifei Zhang: A Graph Reasoning Network for Multi-turn Response Selection via Customized Pre-training. AAAI 2020.
2. Jianxing Yu, Xiaojun Quan, Qinliang Su, Jian Yin: Generating Multi-hop Reasoning Questions to Improve Machine Reading Comprehension. WWW 2020.
3. 陈默, 张昱, 赵长宽等. 支持移动协作学习的异构交互模型建构研究. 中国教育信息化. 2019
4. 闫晟, 徐彬. 程序设计基础课程中结对编程实践与效果评估. 第三届中国计算机实践教育学术会议 CPEC2019.

□ 教学交互生成与自适应优化

- 知识图谱的分段嵌入方法
- 基于网络交互的学习群组与重叠群组发现算法
- 基于问题生成和卷积神经网络的常识问答方法



1. Wentao Xu, Shun Zheng, Liang He, Bin Shao, Jian Yin, Tie-Yan Liu: SEEK: Segmented Embedding of Knowledge Graphs. ACL 2020.
2. 于思淼. 基于网络学习表示的学习群组发现. 硕士学位论文. 2019
3. 一种基于问题生成和卷积神经网络的常识问答方法. 周瑞莹, 梁艺阐, 印鉴. (专利申请)

研究进展3：多主体下个性化精准导学技术

28

重点

□ 教学内容资源多维推荐方法

- 以学习兴趣为导向的冷启动推荐模型
- 基于深度学习的个性化练习导学方法
- 个性化推荐系统技术进展

□ 大规模个性化学习计划自适应优化方法

- BiRNN-DKT深度知识跟踪模型
- 面向课程的多目标自适应学习
- 基于自动编码器的证据感知推理文本生成
- 知识追踪技术的分析

□ 基于领域指导的个性化教学模式生成方法

- 面向个性化导学推荐的课程项模型
- 基于学习迁移的知识追踪方法
- 融合学习与遗忘的深度知识追踪模型
- 基于教师教学风格的学生成绩多维度预测方法
- 基于课堂学习的深度学习机器阅读理解训练方法

成果

1. Yuefeng Du; Tuoyu Yan; Xiaoli Li; Jiafan Zhou; Yang Wang; Jing Shan; A Learning Interests Oriented Model for Cold Start Recommendation, SemiBDMA 2020.
2. 武培文. 基于深度学习的个性化练习导学系统的研究与实现. 硕士学位论文. 2019.
3. 刘君良, 李晓光. 个性化推荐系统技术进展, 计算机科学, 2020.

1. Xu B, Yan S, Yang D. BiRNN-DKT: Transfer Bi-directional LSTM RNN for Knowledge Tracing. WISA2019
2. Jieyue Ma; Xiaoli Li; Xin Zhang; Tingting Liu; Yuefeng Du; Tie Li; Curriculum-Oriented Multi-Goal Agent for Adaptive Learning , SemiBDMA 2020.
3. 刘恒宇, 张天成, 武培文, 于戈. 知识追踪综述. 华东师范大学学报 2019.

1. Yuefeng Du; Angzhi Liu; Xiaoguang Li; Baoyan Song; PS-LDA: A Course Item Model for Tutorial Personalized Recommendation, WISA 2020.
2. Liu Hengyu, Zhang Tiancheng, Li Fan, Gu Yu, Yu Ge. Tracking Knowledge Structures and Proficiencies of Students with Learning Transfer. IEEE Access 2020.
3. 李晓光, 魏思齐, 张昕, 杜岳峰, 于戈. LFKT: 学习与遗忘融合的深度知识追踪模型. 软件学报. 2021.
4. 李嘉伟, 费雪, 赵长宽等. 引入教师教学风格的多维度预测学生成绩的方法. (专利)
5. 印鉴, 梁艺阐, 周瑞莹. 一种基于课堂学习的深度学习机器阅读理解训练方法. (专利)

研究进展4：交互式个性化教学环境原型系统

29

重点

- 基于天河二号平台，研制原型系统 i³Class
- 对现有的 e-learning 系统和 MOOC 系统集成，进行应用系统的开发与部署，形成 i³Class 示范性应用
- 基于课题组所在单位负责的辽宁省资源共享课、辽宁省跨校修读学习平台开展验证工作

成果

- 基于天河二号平台的实验环境（中山大学）

- 基于数据挖掘的教学分析平台（应用中）
- 个性化程序设计在线评测系统（进行中）
- 个性化大学计算机学习平台（进行中）
- 实验实训教学管理平台（校企合作，进行中）

- 面向个体的学习活动自我设计模块（进行中）
- 课程资源与语料库的集成模块（进行中）

研究进展5：应用实例

结合辽宁省普通高等学校精品在线开放课程、辽宁省职业教育精品在线开放课程，及跨校修读学分项目改革实践，开展交互式个性化教学环境构建技术与教学改革探索



研究进展5：应用实例

31

课程 | 辽宁省职业教育精品在线 × +
← → C ▲ 不安全 | courses.lnve.net/courses

四 研究亮点

32

亮点1：面向异构关联数据的一致性规则发现方法

动机：教育资源具有多源异构特征，需要使用规则约束去除资源关系的歧义性，使关系语义上表达一致。

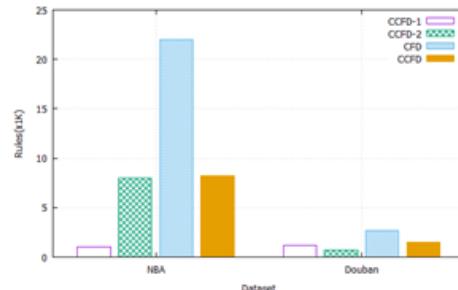
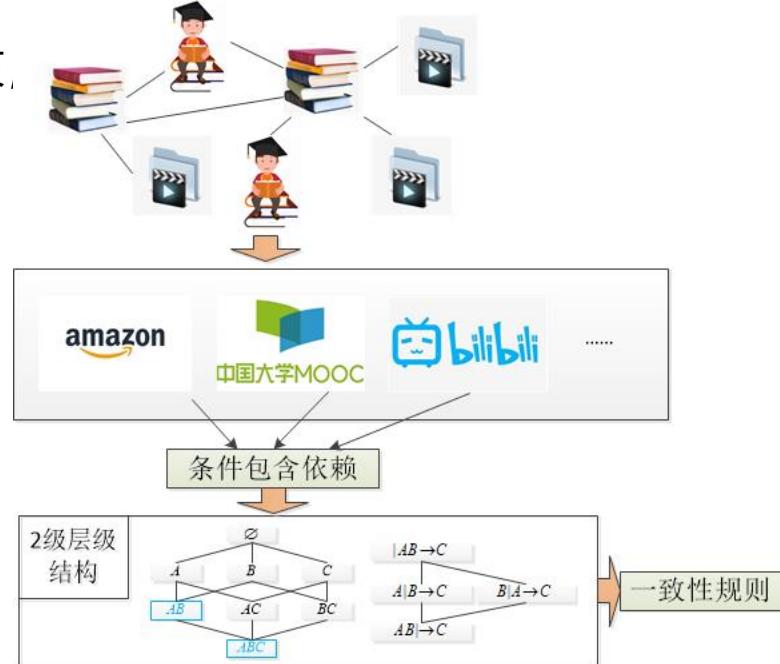
挑战

- 教育资源关系的跨平台发现
- 协作资源之间的潜在关系表达

贡献点

- 提出使用条件包含依赖对异构数据进行匹配的方法，将不同平台的异构资源信息映射到相同的模式中；
- 提出了一种带有协作关系分析的一致性规则，可以捕捉相似资源之间的协作关系；
- 设计了一种基于2级层级结构的规则搜索结构。

$$R_a \oplus_{\Gamma} R_b = \left(\bigcup_{\eta \in \Gamma} X \oplus_{\eta} Y, \bigcup_{\eta \in \Gamma} X_p \oplus_{\eta} Y_p, \right. \\ \left. \bigcap_{\eta \in \Gamma} \text{attr}(R_a) - \text{LHS}(\eta), \bigcap_{\eta \in \Gamma} \text{attr}(R_b) - \text{RHS}(\eta) \right)$$



四 研究亮点

33

亮点2：多主体下智能协作行为分析

动机：

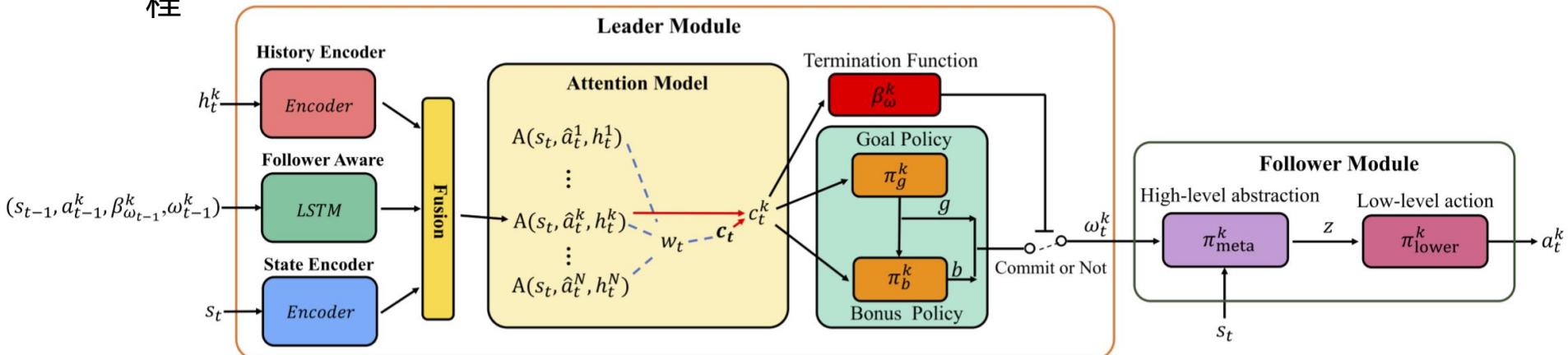
- 小组学习是为了完成一个共同任务，但每位学习者都是有自己的感兴趣的内容和知识结构。

难点：

- 怎样给予奖励来得到更好的长期效果；
- 交互过程太过复杂而难以训练。

主要贡献：

- 提出一种新的基于多代理事件的策略梯度，学习老师的长期策略。
- 提出一种基于动作抽象的策略梯度算法，减少了学生的决策空间，从而加快学生的训练过程



四 研究亮点

亮点3：基于自定义预训练的多轮响应选择图推理网络

动机：

在线教学交互中，在理解对话上下内容的基础上做出合适的回复响应

难点：

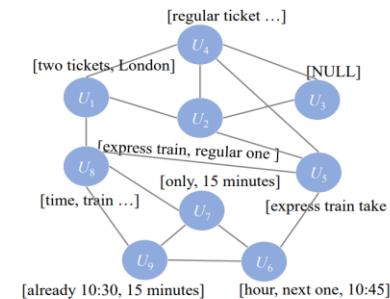
- 日常教学对话具有口语化、上下文省略、多重指代、特殊领域术语等问题；
- 如何对关键信息进行抽取、根据关键信息进行推理。

主要贡献：

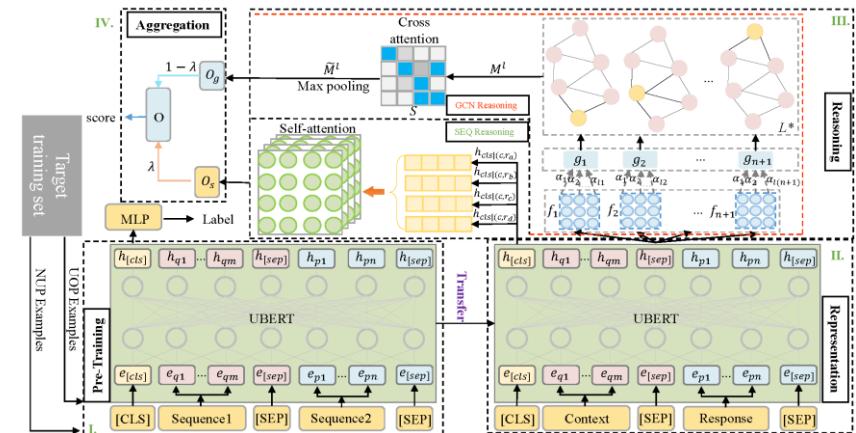
- 根据对话的特点设计了针对性的预训练任务
- 从对话话题的角度将对话文本结构化，构建句子级别的对话图
- 采用序列结构和图结构结合的方式分别从局部和全局的角度进行推理

Turns	Utterance
U1	Good morning , two tickets to london , please .
U2	Express train or regular one ?
U3	Any difference ?
U4	The regular ticket is 80 dollars while price of the express ticket is 10 % higher .
U5	I see , but how long does the express train take ?
U6	Comparatively , it can save you nearly an hour and the next one is leaving at 10:45 .
U7	Oh , we only have 15 minutes to go .
U8	But that 's the time to catch the train . Besides , as long as I go to london earlier , I don 't mind paying a little extra .
U9	Yeah , you 're right . it 's already 10:30. you only have 15 minutes to go .

(a)



(b)

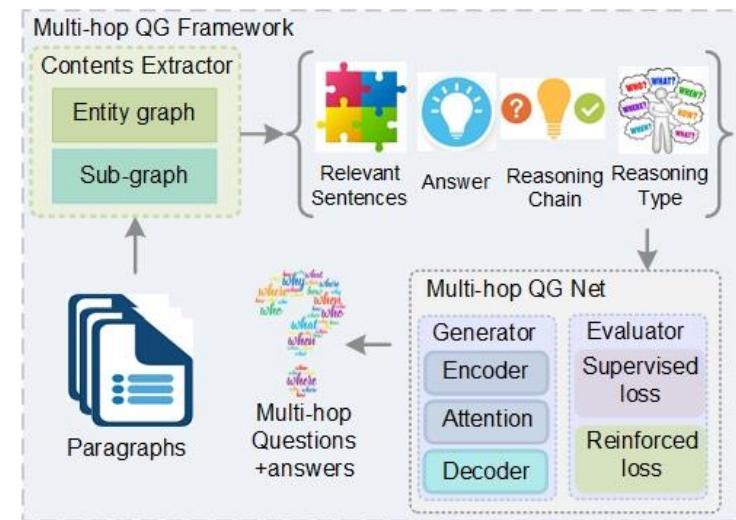
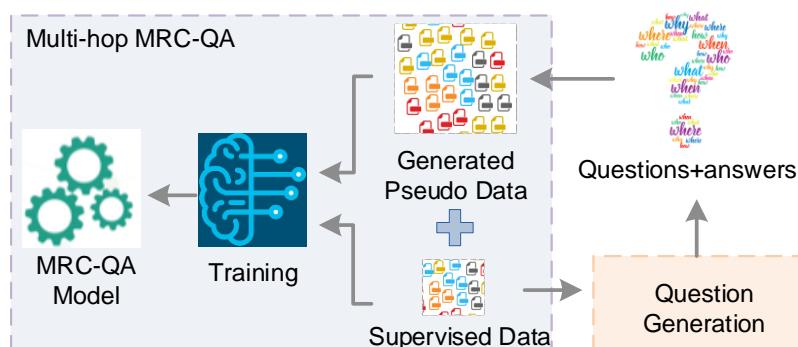


四 研究亮点

35

亮点4：基于数据增强的可推理问答生成

- **动机**：大规模在线教学系统中如何生成大量高质量的考题：考题不但需要能够检验学生对单个知识点的把握能力，而且需要重点考核知识点之间关联的逻辑推理能力
- **挑战**：人工出题非常耗费人力，而自动出题模型不但缺乏可推理能力，还受限于标注数据的规模；如何训练机器标注，如何保证结果是合法且可推理
- **贡献**：提出了一个整体的生成者-评价者网络来形成问题，提升准确率 $>8\%$ ：提问不但需要在句法和语义上合法，提问可推理作答，即Q和A存在多阶推理关联



四 研究亮点

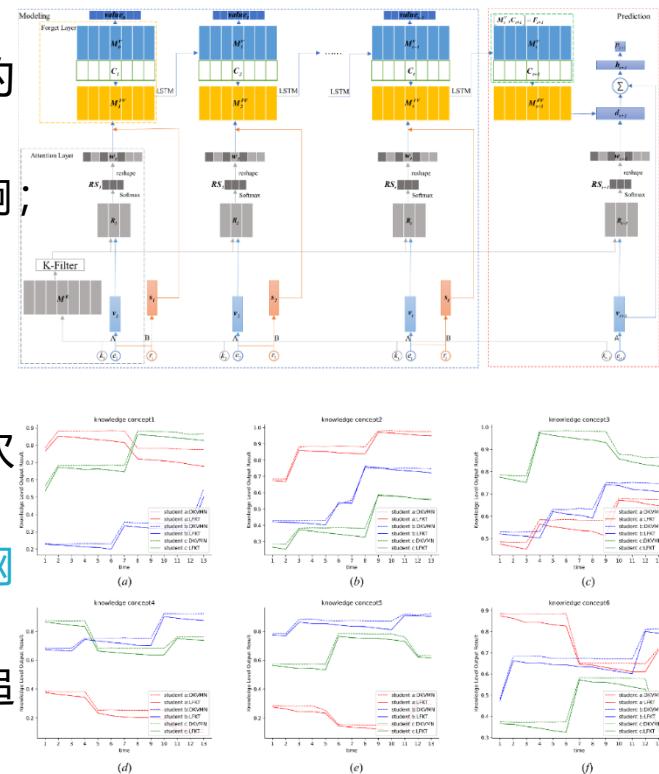
亮点5：LFKT：学习与遗忘融合的深度知识追踪模型

动机：

- 实时追踪学生的知识水平并且预测学生在未来学习中的表现
- 现有系统忽略了学生原本知识水平对于遗忘程度的影响；
- 没有建模学生对知识空间中的各个知识点的遗忘情况。

主要贡献：

- 结合教育心理学，LFKT考虑**四个影响知识遗忘的因素**：学生重复学习知识点的间隔时间、重复学习知识点的次数、顺序学习间隔时间以及学生对于知识点的掌握程度.
- 基于深度神经网络，设计了一个**基于RNN和记忆神经网络的知识追踪神经网络**。
- LFKT可以有效的建模学生的学习行为与遗忘行为,实时追踪学生的知识水平.



五 总结与展望

- 智慧教育是教育学、心理学、信息技术等学科的交叉领域
- 人工智能、大数据、云计算、物联网、区块链等新技术的应用方兴未艾
- 新技术的应用必将推动教育的模式变革和生态重构
- 相关研究：
 - 知识图谱
 - 学习者画像
 - 学习分析和评价
 - 交互技术
 - ...

五 总结与展望

38

□ 下一步工作重点

- 加强教育理论学习，以理论为指导
 - 现代教育学、教育心理学
- 加强教育企业合作，以问题为导向
 - 教育软件公司联合开发
- 发挥超算平台作用，奠定示范基础
 - 天河II号的天河星光容器云平台
- 为示范应用做准备
 - 面向在校学生和社会学习者





谢谢！