

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



认知诊断前沿探索-精准学习状态建模

博士研究生 周瑾洁

2024年04月21日

- 总结反思

- 对基础知识的铺垫较少，接受难度大
- 一些重难点部分需要使用文字解释细节
- 问题引入部分略急，前期演讲过程不太流畅

- 相关内容

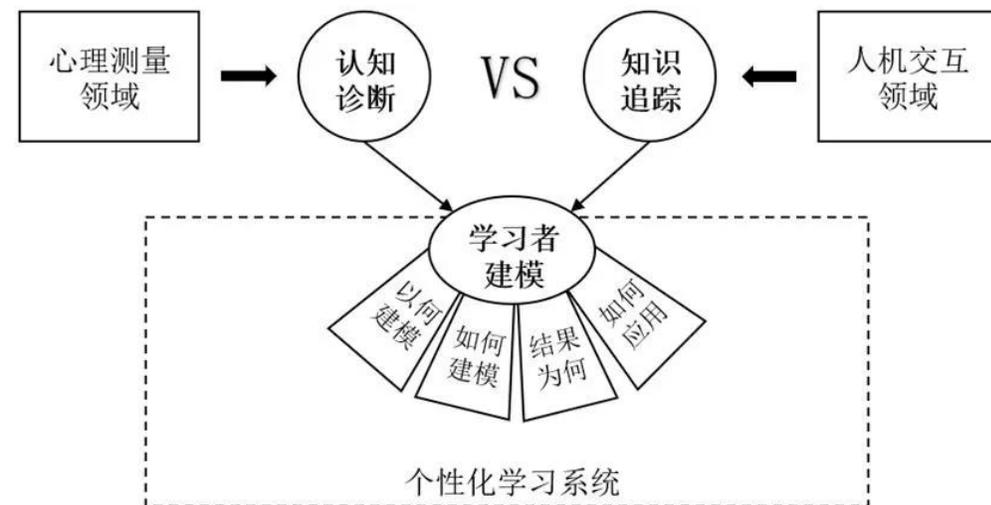
- 2024.04.14 杨晓楠 《**个性化学习**路径推荐—量身打造专属于你的学习Plan》
- 2023.12.17 吴松凌 《强化学生个性的**知识追踪**》
- 2023.09.03 杨晓楠 《**认知诊断**技术及其研究》
- 2023.05.07 周瑾洁 《基于图的**知识追踪**方法研究》
- 2022.03.28 周瑾洁 《从赋能学习到**知识追踪**》

- 预期收获
- 题目内涵解析
- 研究背景与意义
- 研究历史与现状
- 知识基础
- 算法原理
 - ACD
 - GLNC
- 特点总结与工作展望
- 参考文献

- 预期收获
 - 1.了解知识追踪和认知诊断的基本概念和区别
 - 2.熟悉知识追踪和认知诊断的分类及应用场景
 - 3.理解认知诊断前沿技术的应用原理
 - 4.了解认知诊断的发展前景

• 研究目标

- 以**学习者建模**为研究对象，面向智能教学系统（ITS）中认知诊断和知识追踪任务
- 结合深度学习、心流理论、对比学习等技术
- 深入分析学习者的**认知能力**，为教学设计提供更深层次的支持和指导



• 内涵解析

- 认知诊断：研究学习者某一**时间段**学习行为，分析挖掘基于这些行为的**认知状态**（如对所需概念的熟练程度）
- 知识追踪：根据学生的历史学习行为，实时跟踪学生的**知识状态变化**，并预测学生未来的学习表现（**预测学生答对新问题/练习的概率**）

- 研究背景

- 传统教学模式下，由于**时间**和**空间**的限制，教师无法为每一个学生都制定**个性化**学习方案
- 在线学习模式方兴未艾，国内外涌现出大量在线学习系统
- 我国在线教育用户规模已达**4.23**亿，规模空前

- 研究意义

- 深化智能教学系统，致力于**优化**学习者的**个性化学习体验**，具有重要的**理论意义**
- 通过准确预测学习者的表现和深入分析其认知能力，推动教育技术**前沿**发展
- 为教育教学开辟了新的可能性，助力塑造更为丰富、精准的学习**环境**



Junker提出了**DINA**模型，使用人工标注Q矩阵表示练习和概念之间的相关性，并考虑学习者在回答练习时的**猜测**和**失误**因素

Wang等提出神经认知诊断框架NeuralCDM，**首次通过神经网络**来学习学习者与练习之间复杂的交互，使用单调性假设来确保模型可解释性

Su等提出了**GCDM**，通过异构认知图边传播学生的认知状态，并选择性地聚合来自相邻图节点的消息

Su等人提出了一种新颖的全局和局部神经认知（**GLNC**）模型来捕获学生练习交互中的全局和局部特征，结合认知诊断和知识追踪的优势互补

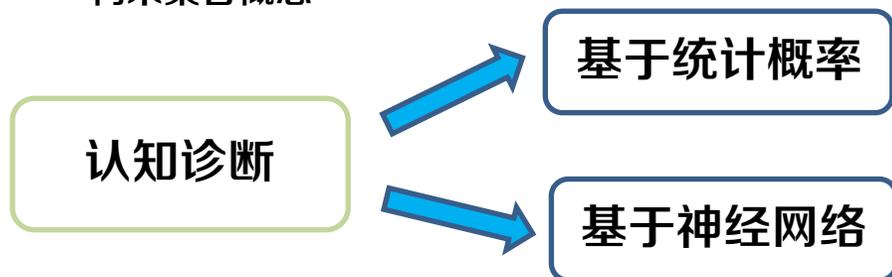


Reckase提出了**MIRT**，将学习者对每个知识概念的熟练程度建模为潜在特征，从而可以诊断学习者对不同知识概念的熟练程度

Wang等提出了**CDGK**，运用神经网络捕捉练习特征、学习者分数和学习者对概念熟练程度之间的相互作用，并将概念转换为图结构来聚合概念

Wang等提出了**KaNCD**，采用基于表示的方法对知识关联进行隐式建模，有效解决认知诊断中的知识覆盖练习

Wang等人提出了一种新颖的情感感知认知诊断（**ACD**）模型，该模型可以通过考虑**情感因素**来有效地诊断学生的知识熟练程度





Corbett提出了贝叶斯知识追踪方法。该方法的核心为**贝叶斯公式**，可解释性强；但模型过于简单，忽略遗忘效应

Piech等**率先**将深度学习用于知识追踪，提出**DKT**，是最经典的深度学习知识追踪方法

Nakagawa等提出了图知识追踪方法。该方法建模了**概念依赖**关系，但忽略了练习本身及练习之间的关系

Huo等提出了**动态学生分类**知识追踪，通过根据学生最近的回答表现定期对学生进行分类，实现个性化预测

Su等人提出了全局和局部神经认知（**GLNC**）模型捕获学生练习交互中全局和局部特征，结合认知诊断和知识追踪的优势互补

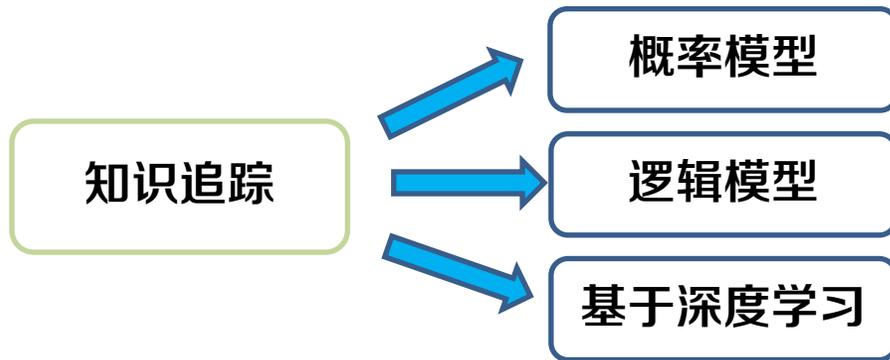


Rasch将**项目反应理论**用于知识追踪，通过考虑练习的难度水平和学生的能力水平来预测学生正确回答一个练习的概率

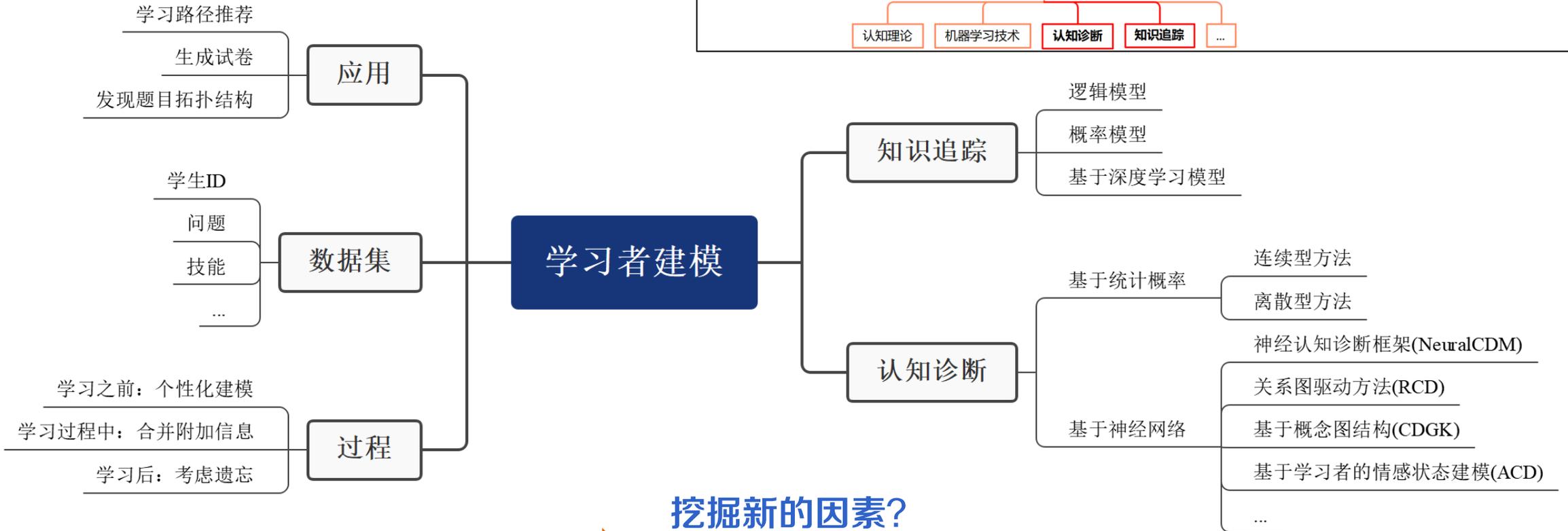
Zhang等提出了动态键值记忆网络，分别用**键矩阵**存储概念并表示，**值矩阵**存储学生对每个概念的掌握程度，提高可解释性

Pandey等提出了自注意力知识追踪，使用**Transformer**模型考虑学习历史中相关练习来预测学生回答下一个练习的表现，处理数据稀疏问题

Gan等提出了一种知识结构增强图知识追踪（**KSGKT**）的KT模型，该模型根据学习者在在线学习系统中的锻炼日志来追踪学习者知识熟练程度的演变



• 认知诊断和知识追踪研究分类



前沿方法如何精确建模?



挖掘新的因素?
与知识追踪结合?
练习标签化?

• 认知诊断

– 目标：基于学生某一**时间段**的学习行为，分析其**认知状态**

- 学生集合, $S = \{s_1, s_2, s_3, \dots, s_N\}$
- 练习集合, $E = \{q_1, q_2, q_3, \dots, q_M\}$
- 概念集合, $K = \{k_1, k_2, k_3, \dots, k_K\}$
- 答题记录, $R = \{(s, q, r) | s \in S, q \in Q\}$
- Q矩阵：反映**练习与概念**间的映射关系

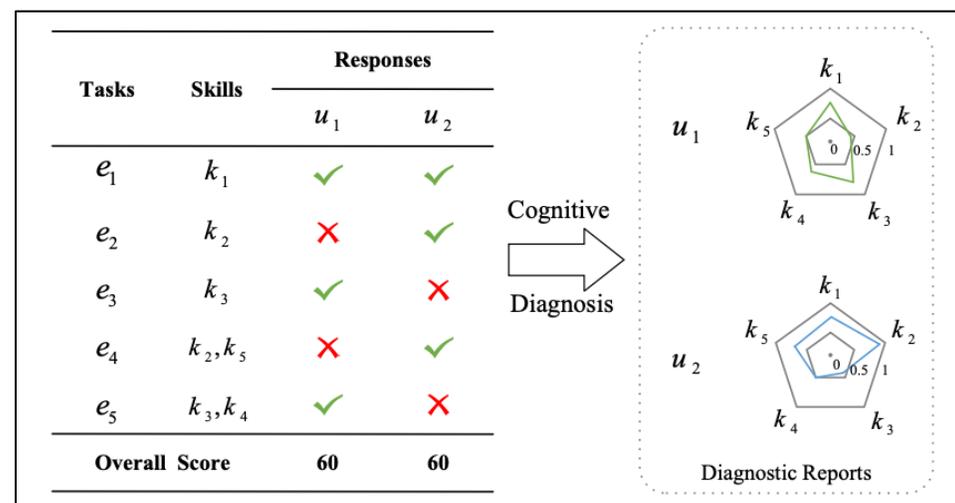
$$Q = \{Q_{i,j} \in \{0,1\}\}_{M \times K}$$

– 给定学生**答题记录**和**Q矩阵**，通过答题预测过程获取学生对各个概念的**掌握程度**

学生对知识概念的掌握程度

Q矩阵

	Questions					
	q1	q2	q3	q4	q5	q6
con1	1	0	0	0	0	1
con2	1	1	0	1	0	0
con3	1	1	1	0	0	0
con4	1	0	1	0	0	0

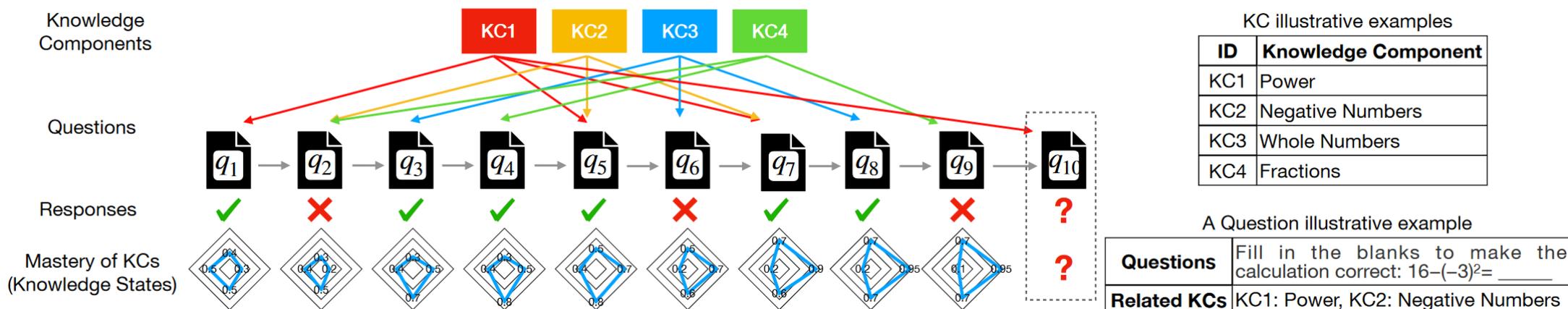


知识追踪

– 目标：精确预测学生答对新练习的能力

- 基于学习历史-问答序列 $X_t = \{x_1, x_2, \dots, x_t\}$ ，模型提取跟踪学生隐式知识状态
- $x_t = (q_t, r_t)$ ，学生在时刻 t 回答了 q_t ，回答为 $r_t \in (0, 1)$
- 模型预测学生答对下一个练习的概率， $P(a_{t+1} = correct | q_{t+1}, X_t)$

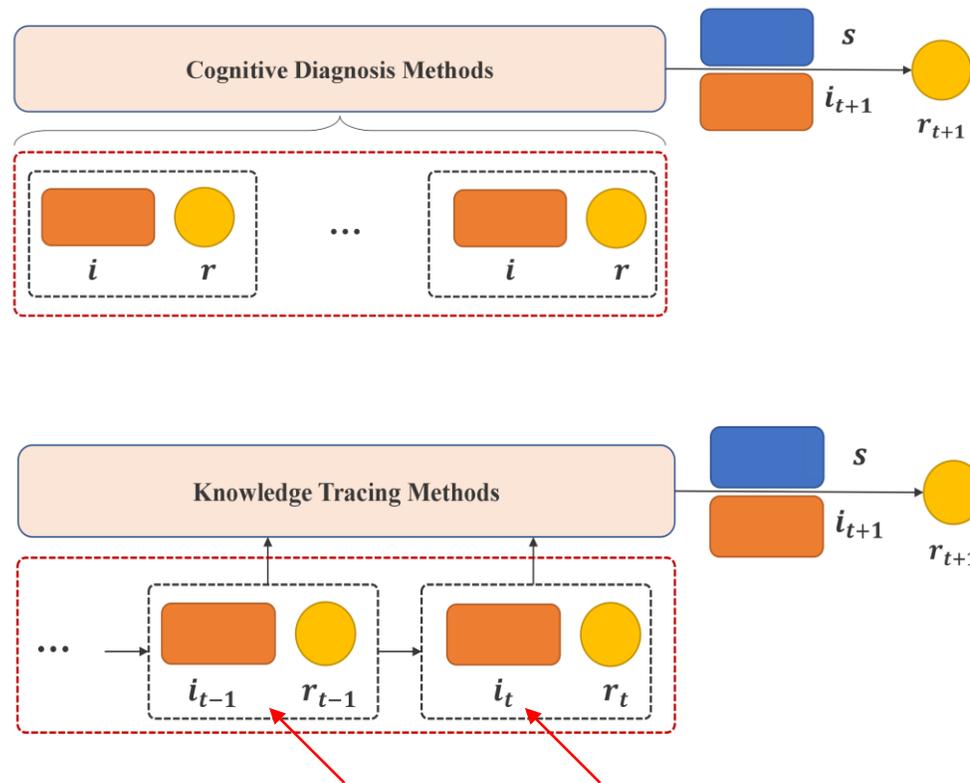
– 基于学生历史交互进行建模，预测学生答对新练习概率



设计算法，精准预测

• 认知诊断与知识追踪区别

比较维度	认知诊断 (CD)	知识追踪 (KT)
T 目标	评估学生对知识点 (概念) 的掌握程度	预测学生在下一个练习的回答正确率
I 输入	学习记录、Q矩阵	学习记录、Q矩阵
O 输出	学生知识点掌握程度 预测答题情况	预测学生答题情况
模型参数	可解释性	黑盒参数
应用场景	自适应测试 、个性化推荐	学习成绩预测、个性化精准推荐
领域出发点	心理测量	人机交互
特点	全局静态评估 , 结果精细化、多样化	局部动态预测 , 结果时序化



全局静态和局部动态



• 认知水平研究范式

– DINA (deterministic input, noisy “and” gate)

- 将学生描述成一个多维的**知识点掌握向量**，根据学生实际作答结果进行诊断

– 学生 i 在练习 j 的作答情况

$$n_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} = \begin{cases} 0 \\ 1 \end{cases}$$

– 概率预测函数

$$P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = g_j^{1-n_{ij}}(1 - s_j)^{n_{ij}}$$

- 失误: s_j ，学生在掌握练习 j 所有概念情况下，做错的概率
- 猜测: g_j ，学生在不完全掌握练习 j 所有概念情况下，猜对的概率

交互函数由人工设计、函数模拟能力有限、知识点过多导致计算复杂

基于神经网络的认知诊断框架-NCD

- 学生建模: 学生知识熟练度向量: $h^s = \text{sigmoid}(x^s \times A)$
- 练习建模
 - 知识点相关度向量

$$Q_e = x^e \times Q$$

- 知识点难度向量

$$h^{diff} = \text{sigmoid}(x^e \times B)$$

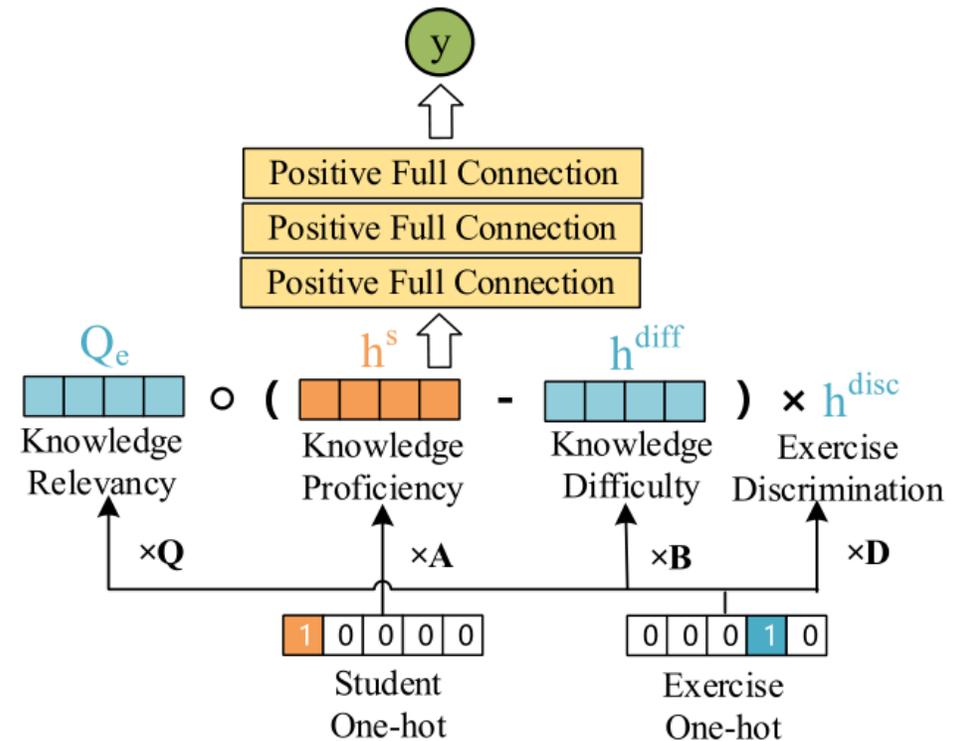
- 练习区分度向量

$$h^{disc} = \text{sigmoid}(x^e \times D)$$

- 交互函数预测

$$f_0 = Q^e \circ (h^s - h^{diff}) \times h^{disc}$$

$$y^* = \text{sigmoid}(W_n \times f_{n-1} + b_n)$$





**Affect-aware Cognitive Diagnosis
(ACD)**

T	目标	分析学生在学习中的认知状态和情感状态间关系，以全面认知诊断
I	输入	学生学习交互数据（答题情况、练习编号、概念编号），教育数据集*3 Junyi（2100万交互），ASSIST17（94万交互），ASSIST12（270万交互）
P	处理	1.计算认知诊断基础向量和学生潜在情感向量 2.学生情感感知模块预测学生情感 3.NCD框架预测学生回答的正确率 4.情感感知认知诊断模块 结合 失误猜测参数和NCD结果，最终预测
O	输出	试题回答结果（正、误）、特定知识概念掌握程度（范围(0,1)）

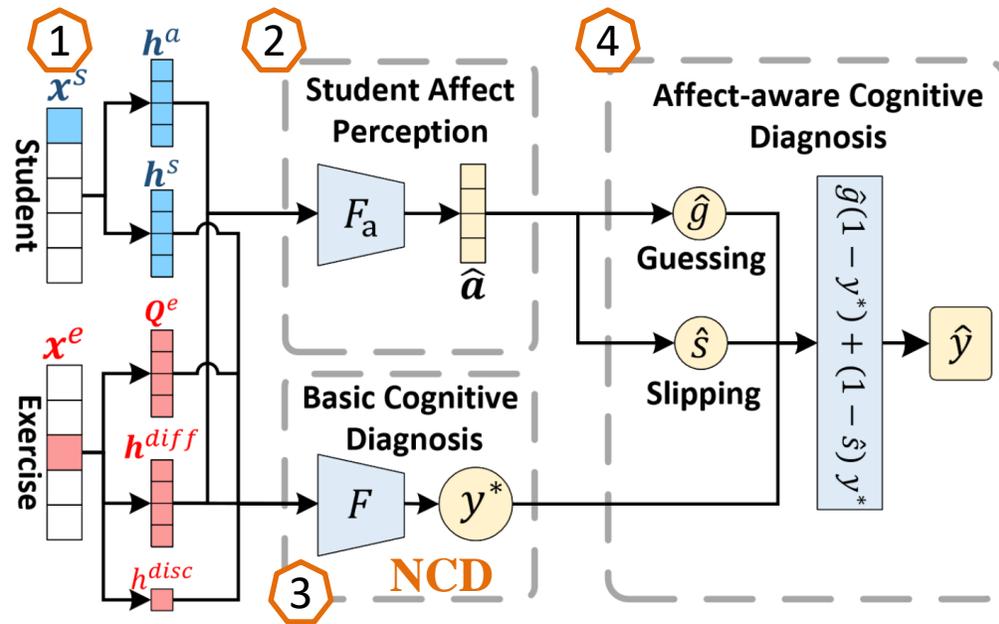
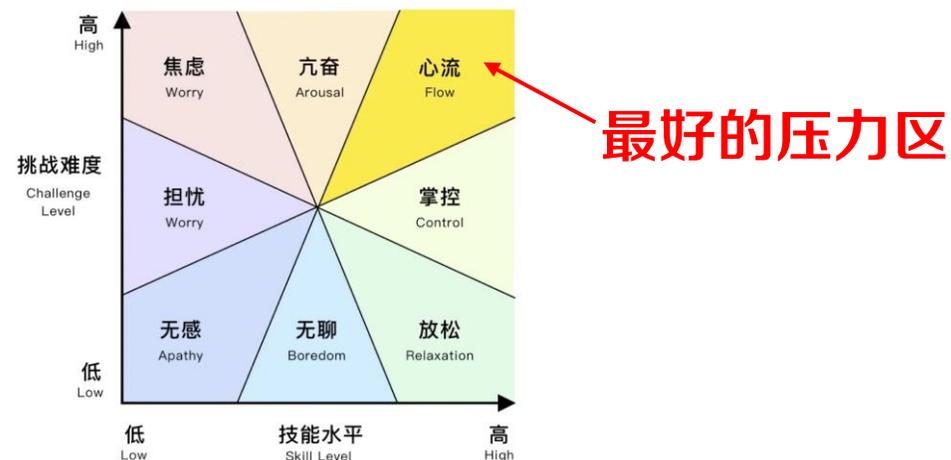
P	问题	未考虑学生在学习过程中情感状态
C	条件	基于 NCD 为基础认知诊断框架 情感因素和 DINA 中的 失误 和 猜测 参数强相关
D	难点	分析情感状态和回答练习的关系；挖掘无情感标签数据集的情感因素
L	水平	AAAI 2024（CCF-A）

核心思想

- 心流理论，情感取决于挑战难度和技能水平
- 挖掘学生认知状态和情感状态间关系
- 即插即用情感感知模块，兼容有监督或无监督的学习环境

算法步骤

- 1. 计算认知诊断基础向量和学生潜在情感向量
- 2. 学生情感感知模块，预测个性化情感
- 3. NCD框架，预测认知状态
- 4. 情感感知认知诊断
 - 推断失误和猜测参数
 - 结合DINA 和NCD诊断结果，最终预测



计算认知诊断基础向量和学生潜在情感向量

– 基础定义

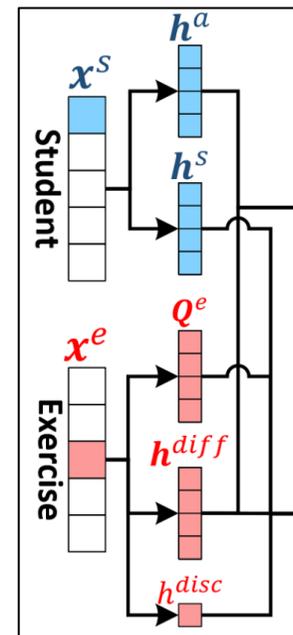
- 学生集合: $S = \{s_1, s_2, s_3, \dots, s_N\}$, 练习集合, $E = \{q_1, q_2, q_3, \dots, q_M\}$
- 概念集合: $K = \{k_1, k_2, k_3, \dots, k_K\}$, Q矩阵: $Q = \{Q_{i,j} \in \{0,1\}\}_{M \times K}$

– 学生建模

- 学生对知识概念的掌握程度: $h^s = \text{sigmoid}(x^s \times S)$
- 学生潜在情感向量: $h^a = \text{sigmoid}(x^s \times A)$

– 练习建模

- 知识点相关度向量: $Q^e = x^e \times Q$, $Q^e \in \{0,1\}^{1 \times K}$, $x^e \in \{0,1\}^{1 \times N}$
- 知识点难度向量: $h^{diff} = \text{sigmoid}(x^e \times E)$, $h^{diff} \in (0,1)^{1 \times K}$
- 练习区分度向量: $h^{disc} = \text{sigmoid}(x^e \times D)$, $h^{disc} \in (0,1)$



学生情感感知模块

– 核心思想：当学生解决练习时，受到**情感**和**练习难度**的**共同影响**

• **沮丧、困惑、专注、无聊**

– 学生在练习上的个性化情感

• 预测学生情感向量

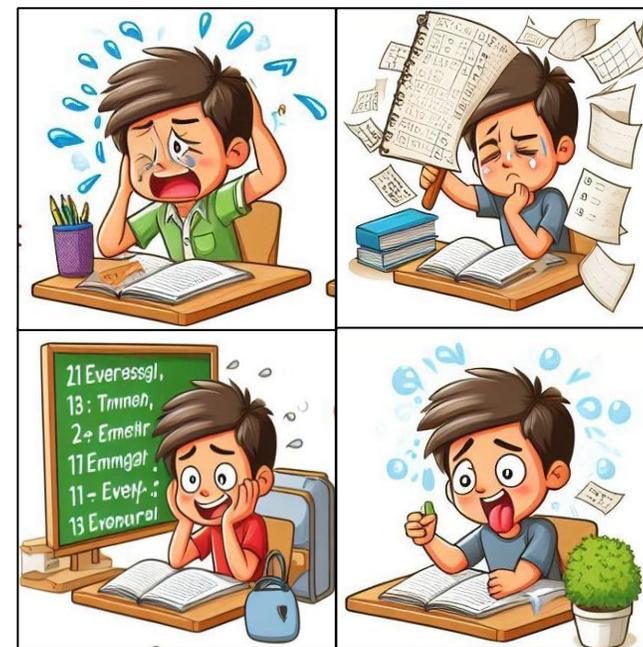
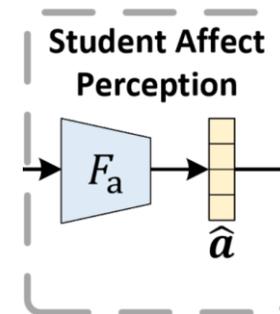
$$\hat{a} = \text{sigmoid}(W_a \times [h^a, h^{diff}]^T + b_a), \quad \hat{a} \in (0,1)^Z$$

• 均方误差（MSE）损失函数训练

$$\mathcal{L}_a = \sum_i \|\hat{a}^i - a_{gt}^i\|^2$$

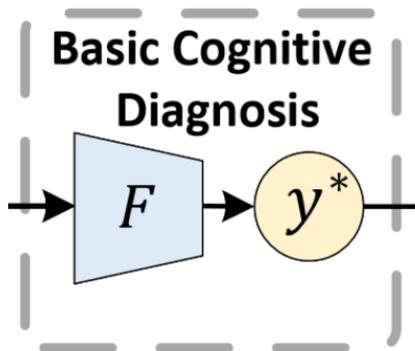
– a_{gt}^i 表示第 i 个学生练习交互的带注释的情感向量

– 基于学生情感分布，推断**失误**和**猜测**参数

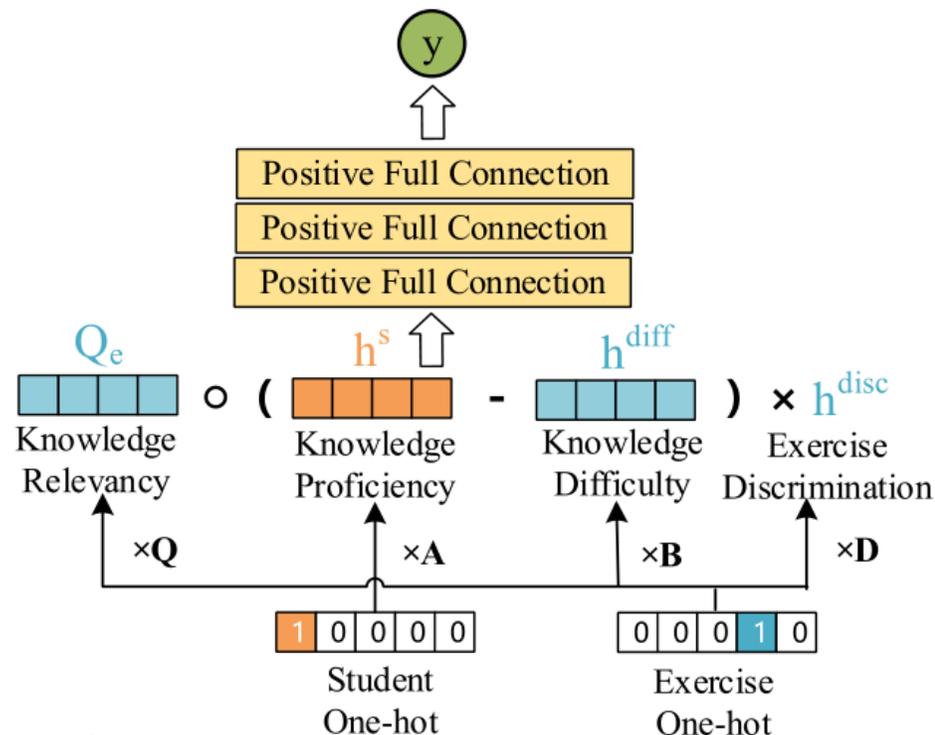


情感和难度共同决定学生回答练习的正确率

基于神经网络的认知诊断框架-NCD



=



- 交互函数预测

$$f_0 = Q^e \circ (h^s - h^{diff}) \times h^{disc}$$

$$y^* = \text{sigmoid}(W_n \times f_{n-1} + b_n)$$

- 所有权重矩阵为正值以满足**单调性假设**

- 单调性假设：学生答对练习的概率随学生对任意知识点熟练度单调递增

掌握知识越多，答对概率越高

情感感知认知诊断

- 基于学生情感分布，推断**失误**和**猜测**参数

$$\hat{g} = \text{sigmoid}(W_g \times \hat{a} + b_g), \hat{g} \in (0,1)$$

$$\hat{s} = \text{sigmoid}(W_s \times \hat{a} + b_s), \hat{s} \in (0,1)$$

- 采用 DINA 诊断公式来推断学生情感反应分数

$$\hat{y} = \hat{g}(1 - \gamma^*) + (1 - \hat{s})\gamma^*$$

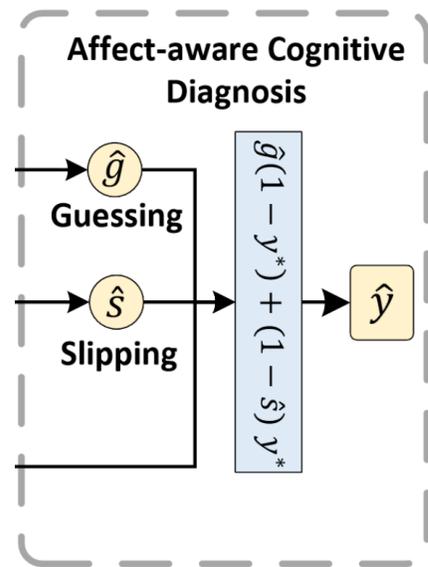
- 二元交叉熵损失，答案标签 $r \in (0,1)$

$$\mathcal{L}_{\text{CDM}} = - \sum_i (r_i \log \hat{y}_i + (1 - r_i) \log(1 - \hat{y}_i))$$

- 结合情感感知认知诊断模块和NCD诊断结果，采用**联合**损失函数优化模型

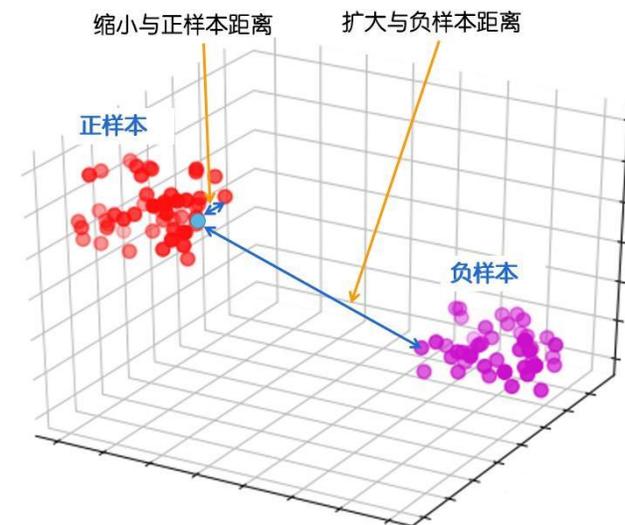
$$\mathcal{L} = \mathcal{L}_{\text{CDM}} + \lambda \mathcal{L}_a$$

个性化情感向量推断参数，参与认知诊断任务



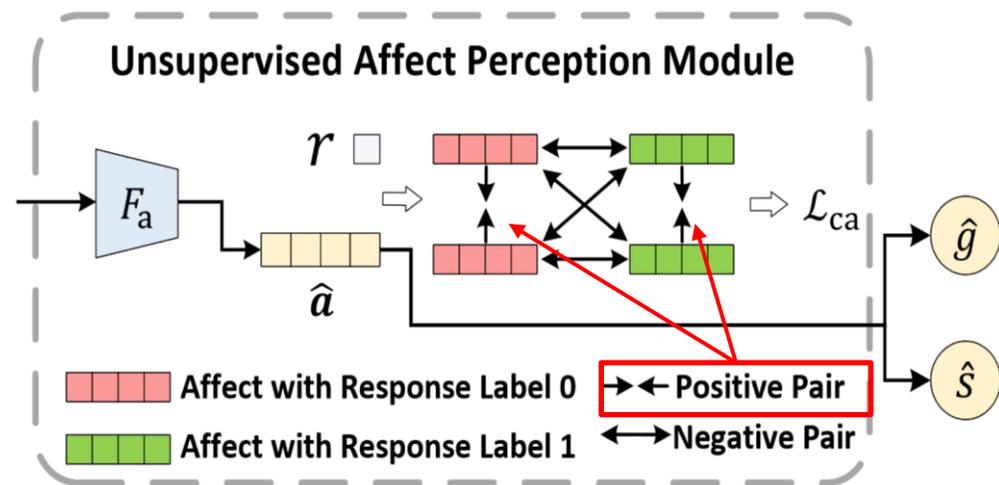
无监督对比 ACD 模型 (CACD)

- 具备学生情感标签的数据集**较少**
 - 难以优化学生情感感知模块
- 设计**对比学习**情感感知模块，优化学生情感向量
 - 正样本：学生-练习交互结果对应学生**相同**情感状态
 - 负样本：学生-练习交互结果对应学生**不同**情感状态



$$\mathcal{L}_{ca} = - \sum_i \frac{1}{|\mathcal{P}_i|} \sum_{j \in \mathcal{P}_i} \log \frac{\exp(\text{sim}(\hat{a}^i, \hat{a}^j) / \tau)}{\sum_{k \in \mathcal{N}_i} \exp(\text{sim}(\hat{a}^i, \hat{a}^k) / \tau)}$$

- \mathcal{P}_i : 第 i 个学生练习交互构建的**正**对集合
- \mathcal{N}_i : 第 i 次学生练习交互构建的**负**对集合
- τ : 温度系数



情感积极促进回答正确率

- 数据集*3
 - Junyi (超过2100万学习记录) , 不包含情感信息
 - ASSIST17 (超过94万学习记录) , 包含无聊、集中、困惑和沮丧情感
 - ASSIST12 (超过270万学习记录) , 包含无聊、集中、困惑和沮丧情感
 - 删除答题记录少于15条的学生
- 评价指标
 - AUC、ACC、RMSE
- 对比方法
 - 经典: DINA(2009)、IRT(1952)、MIRT(2009)、NCD(2020)
 - 先进: RCD(2021)、SCD(2023)



- 有情感标签数据集
 - 所有结合ACD方法都优于各自的基线，对早期方法改进显著
 - RCD和SCD提升较小，图神经网络虽增强性能，但削弱特征判别力
- 无情感标签数据集
 - 即使没有情感标签，CACD仍提升较大，情感属性有效

Datasets		ASSIST17			ASSIST12		
Models		ACC	RMSE	AUC	ACC	RMSE	AUC
DINA	baseline	64.84±0.09	46.66±0.02	69.64±0.06	71.45±0.07	43.80±0.04	69.49±0.11
	ACD	71.15±0.26	43.62±0.18	77.72±0.29	73.76±0.08	42.12±0.05	74.21±0.06
IRT	baseline	65.96±0.10	46.53±0.03	72.37±0.07	73.11±0.04	42.67±0.01	72.67±0.03
	ACD	71.64±0.23	43.25±0.10	78.55±0.22	74.26±0.07	41.71±0.02	75.44±0.07
MIRT	baseline	68.17±0.02	46.48±0.03	74.13±0.00	73.81±0.00	44.44±0.00	72.58±0.00
	ACD	72.43±0.13	42.78±0.07	79.46±0.13	74.38±0.07	41.63±0.06	75.64±0.11
NCD	baseline	69.21±0.87	45.13±0.60	75.34±1.01	74.23±0.05	41.95±0.01	74.78±0.04
	ACD	72.42±0.10	42.84±0.07	79.49±0.10	74.98±0.05	41.37±0.06	76.02±0.16
RCD	baseline	71.55±0.15	43.39±0.10	78.10±0.03	74.49±0.04	41.75±0.03	75.22±0.05
	ACD	72.31±0.10	42.84±0.05	79.27±0.08	74.61±0.02	41.43±0.00	76.19±0.01
SCD	baseline	71.59±0.10	43.35±0.05	78.19±0.01	74.64±0.03	41.48±0.02	75.94±0.04
	ACD	72.69±0.05	42.79±0.02	79.40±0.07	74.70±0.01	41.37±0.00	76.10±0.02

有情感标签数据集

Dataset		junyi		
Models		ACC	RMSE	AUC
DINA	baseline	74.22	41.76	78.72
	CACD	76.96	40.05	82.18
IRT	baseline	67.60	42.68	77.50
	CACD	77.17	39.76	82.64
MIRT	baseline	75.13	41.17	79.89
	CACD	77.27	39.73	82.64
NCD	baseline	74.43	41.72	79.09
	CACD	77.35	39.70	82.73
RCD	baseline	77.16	39.63	82.62
	CACD	77.42	39.56	82.94
SCD	baseline	77.30	39.61	82.77
	CACD	77.45	39.59	82.90

无情感标签数据集 24

验证实验-情感感知的可靠性 (ASSIST17)

– 对比方法:

- NCD不涉及情感预测，使用**随机**情感
- ACD-w/o-L 变体，没有情感预测Loss， $\mathcal{L} = \mathcal{L}_{CDM} + \lambda\mathcal{L}_a$ ， $\lambda = 0$
- Oracle变体，直接从情感标签中提取个性化情感

Models	Student Performance			Affect	
	ACC	RMSE	AUC	RMSE	MAE
NCD	69.21	45.13	75.34	45.34	36.39
ACD-w/o-L	71.54	43.37	78.40	42.48	33.12
CACD	71.76	43.25	78.54	32.41	23.28
ACD	<u>72.42</u>	<u>42.84</u>	<u>79.49</u>	<u>22.10</u>	<u>14.64</u>
Oracle	79.84	40.27	80.35	0	0

– 实验结论

- 情感预测的准确性**越高**，认知诊断结果越准确，情感预测和认知诊断的**结合有效**
- 与ACD-w/o-L相比，CACD在学生成绩预测方面提升较小，但CACD在**情感预测**方面提升**较大**，可解释性强



- 学生情感感知的消融研究

- λ 是方程中的权衡参数

- 选取值从0.001到10

$$\mathcal{L} = \mathcal{L}_{\text{CDM}} + \lambda \mathcal{L}_a$$

- $\lambda=1$, 实现最佳性能

- 学生的潜在情感特征和练习难度相关度

- 仅考虑练习难度

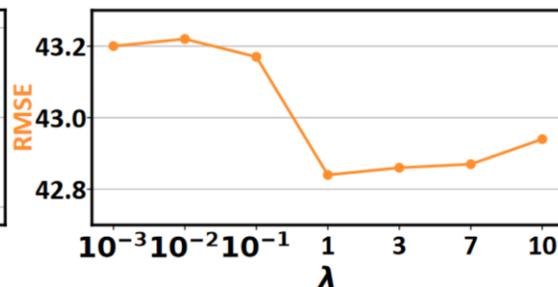
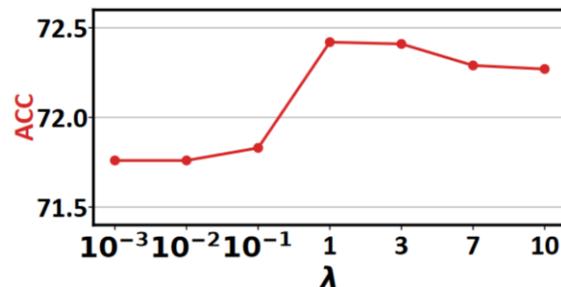
- ACD与DINA类似, 性能优于仅考虑学生情感特征结果

- 表明不同学生在尝试相同难度练习时可能会经历相似的情感

- 结合两个因素

- 性能显著提高

- 表明学生在回答练习时的情感与学生情感特征和练习难度同时相关



student affect conception		ACD	RMSE	AUC
student	exercise			
✓		63.54	47.19	67.83
	✓	65.51	46.35	70.59
✓	✓	72.42	42.84	79.27

情感和学生回答结果的相关性

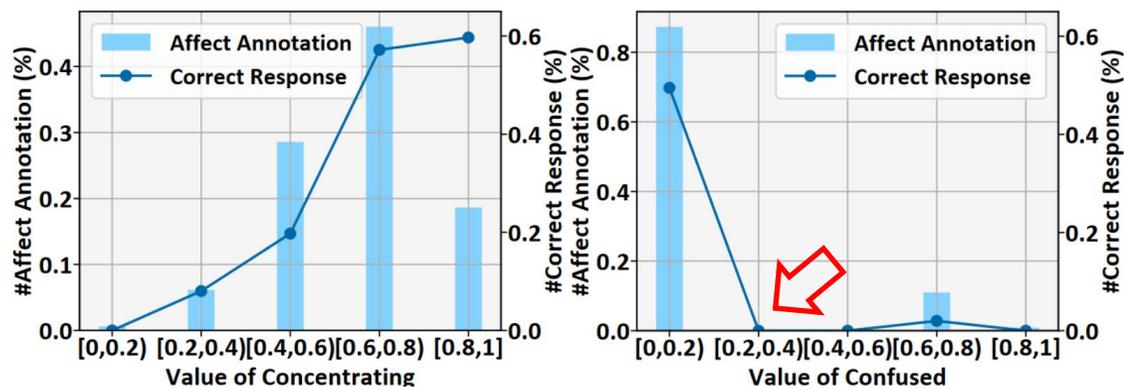
– 不同专注力和困惑影响下，学生正确回答练习概率

- 专注力越高，学生正确回答练习概率高

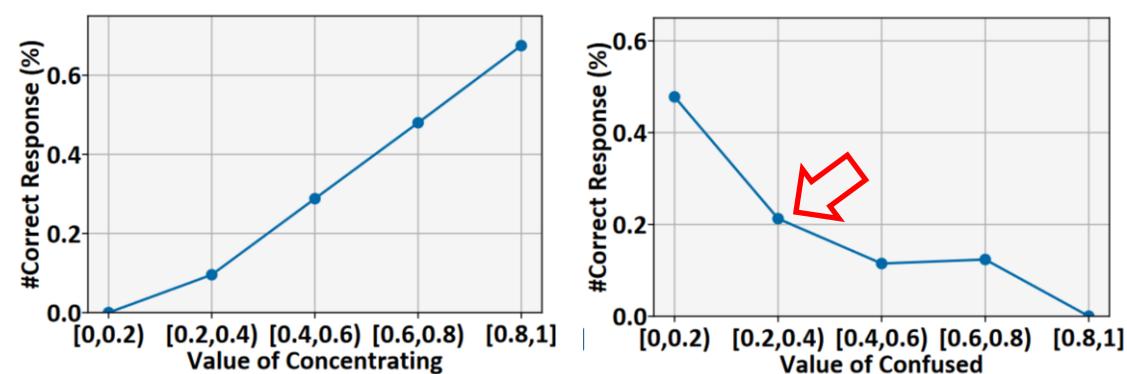
- 困惑感越强，学生错误回答练习概率高

– 预测的情感分布与真实情况之间趋势相似，但存在差异

- 原因：响应受到各种情感和学生能力的组合的影响，而不是单一情感的影响



ASSIST17的统计结果



ASSIST17的预测结果

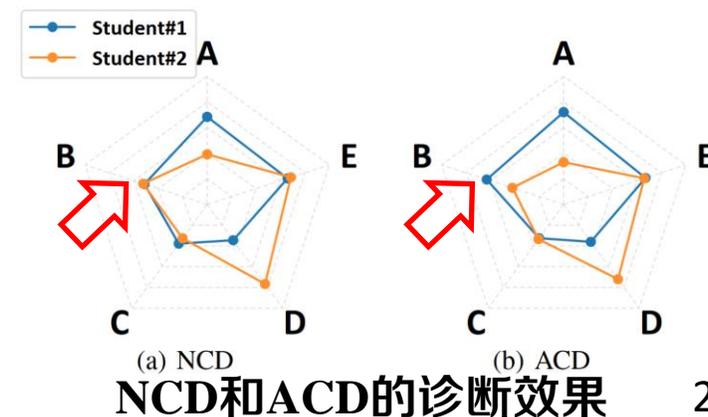
- 实验目的：更全面分析情感的作用
 - 学生的情感与ACD对失误和猜测概率的预测
 - 学生1**失误**：#3回答错误，掌握概念B（#2正确）
 - 学生2**猜测**：#3回答正确，未掌握概念B（#2错误）
 - 实验结论
 - 学生1比学生2对概念B的掌握程度更高
- NCD和ACD的诊断效果展示
 - NCD
 - 两名学生对概念 B 的熟练程度**相似**
 - 没有提供可解释的诊断结果
 - ACD
 - 学生1比学生2对概念B的掌握程度更高
 - 效果与上述**结论一致**，可解释更强

Exercise	#1	#2	#3	#4	#5	#6
Concept	A	B	B	C	D	E
Student#1	✓	✓	✗	✗	✗	✓
Student#2	✗	✗	✓	✗	✓	✓

2名学生在ASSIST12中的答题日志

	Student#1		Student#2	
	Ex-#2	Ex-#3	Ex-#2	Ex-#3
Frustrated	0.3	0.3	0.3	0.5
Confused	0	0	0	0.6
Concentrating	0.7	0.1	0.7	0.7
Bored	0.4	0.7	0.1	0.3
guessing	-	-	-	✓
slipping	-	✓	-	-

情感与ACD对失误和滑倒概率的预测





Global and local neural cognitive modeling for student performance prediction (GLNC)

T	目标	结合知识追踪的知识状态和认知诊断的认知状态，更加全面、精确地评估学生的认知状态
I	输入	学生学习交互数据（答题情况、练习编号、概念编号），教育数据集*3 TAL2023（550万交互），EdNet（900万交互），Eedi2020（1000万交互）
P	处理	1.根据学生先前学习交互，建模其 全局 认知状态 2.自注意力编码器学习最近学生-练习交互关系，提取 局部动态 认知 3. 结合 全局认知状态和局部认知动态
O	输出	试题回答结果（正、误）、特定知识概念掌握程度（范围(0,1)）

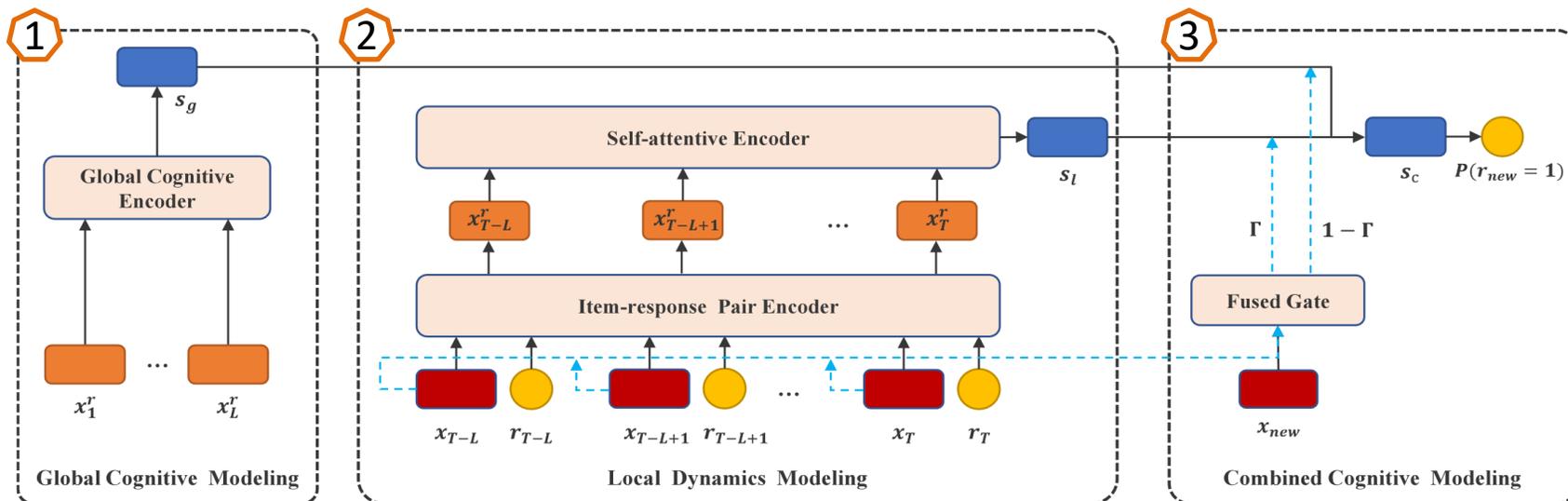
P	问题	认知诊断模型假设学生认知状态是静态的， 无法动态 衡量 知识追踪对交互序列非常 敏感 ，模型 不稳定
C	条件	知识追踪阶段计算学生最近的30次互动练习
D	难点	如何较好地融合知识追踪的知识状态和认知诊断的认知状态
L	水平	Expert Systems With Applications 2024（SCI一区）

• 核心思想

- 综合考虑全局静态特征和局部动态特征

• 算法步骤

- 全局认知建模，学习学生的**全局**认知状态
- 局部动态建模，自注意力编码器提取**局部**动态
- 自适应组合认知建模，**结合**全局认知状态和局部认知动态，进行预测



全局认知建模，学习学生的全局认知状态

– 基础定义

- 学生集合: $\mathcal{S} = \{s_1, s_2, s_3, \dots, s_L\}$, 练习集合: $\mathcal{I} = \{i_1, i_2, i_3, \dots, i_J\}$
- 概念集合: $\mathcal{K} = \{k_1, k_2, k_3, \dots, k_M\}$, Q矩阵: $Q = \{Q_{j,m} \in \{0,1\}\}_{J \times M}$

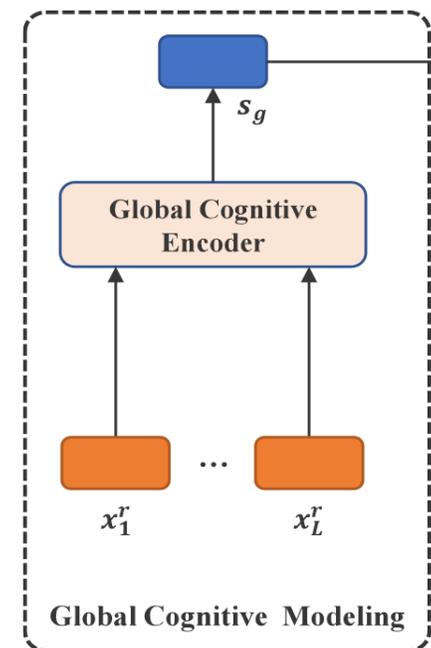
– 学生建模

- 对学生 s , 随机初始化一个全局认知向量 s_g , 表征全局认知状态
- MLP结合练习嵌入 i_j 和概念嵌入 k_{ij} , 增强练习表示

$$x_j = W_1^T (i_j \oplus k_{ij}) + b_1$$

- 设计正确层和错误层, 区分不同效果的交互向量

$$x_j^r = \begin{cases} W_r^T (x_j) + b_r, & \text{if } r_{ij} = 1 \\ W_w^T (x_j) + b_w, & \text{if } r_{ij} = 0 \end{cases}$$



局部动态建模

– 自注意力编码器，捕获学生局部动态认知状态并学习学生-练习交互间依赖关系

• 自注意力注意层，**缩放点积注意力**

$$w_{ll'} = \text{softmax} \left(\frac{q_{l,j}^{rT} k_{l',j}^r}{\sqrt{d_x}} \right), \text{ 缩放因子为 } \frac{1}{\sqrt{d_x}}$$

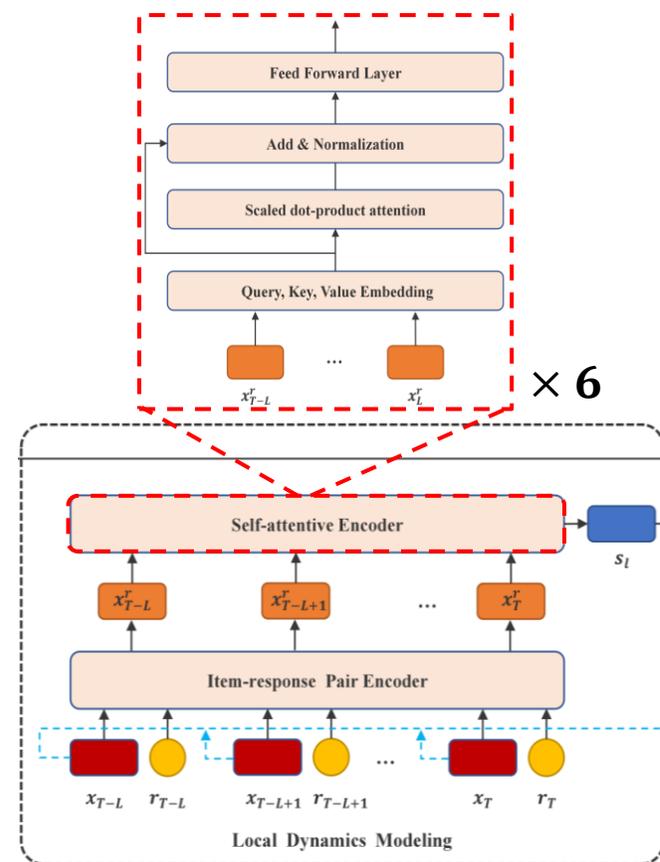
$$a_j^r = \sum_{l'=T-L}^T w_{ll'} v_{l',j}^r$$

• 归一化层， $\hat{a}_j^r = \text{LayerNormalization}(x_j^r + a_j^r)$

• 前馈层， $\tilde{a}_j^r = \text{Relu}(W_2^T \hat{a}_j^r + b_2) + \hat{a}_j^r$

• 全局平均池化和MLP，局部认知向量

$$s_l = W_3^T \frac{\sum \tilde{a}_j^r}{L} + b_3$$



- 自适应组合认知建模，结合全局认知状态和局部认知动态

- 结合练习和知识概念嵌入： $x_{new} = W_1^T (i_{new} \oplus k_{i_{new}}) + b_1$

- 计算学生最近的30次互动练习与新练习间相似性

$$att = \sigma\left(\frac{\sum_{l=T-L}^T x_l^T x_{new}}{L}\right)$$

- 设计融合门

$$s_c = \Gamma \cdot s_l + (1 - \Gamma) \cdot s_g$$

$$\Gamma = \sigma(W_4^T (att) + b_4)$$

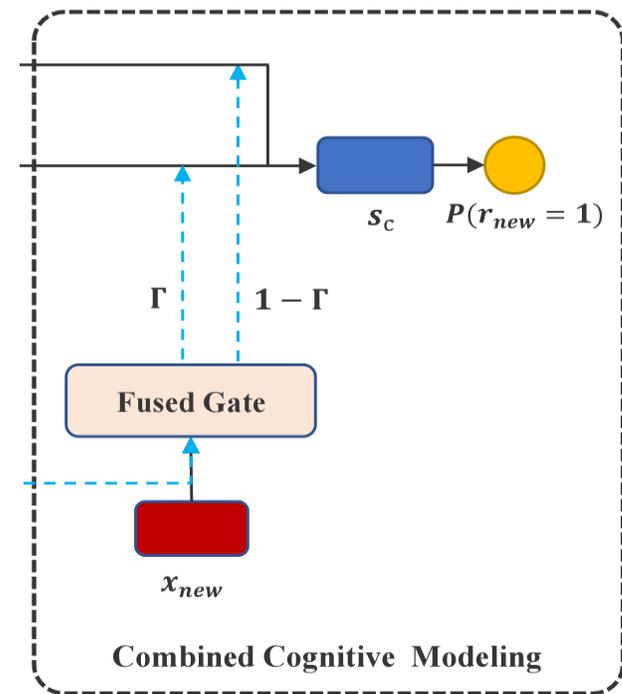
- 认知和练习向量内积，模拟学生回答练习决策过程

$$P(r_{new} = 1) = \sigma(\sum(s_c \cdot x_{new}))$$

- 二元交叉熵损失函数训练模型

$$\mathbb{L} = -\sum(p \log r + (1 - p) \log(1 - r)) + \lambda_{\Theta} \|\Theta\|^2$$

- Θ : GLNC 的所有参数, λ_{Θ} : 正则化超参数



- 问题

- 1.与先进的认知诊断和知识追踪方法相比，GLNC 性能如何？
- 2.GLNC 中的不同元素（例如全局认知建模、局部动态建模、自适应组合和知识概念）如何影响其性能？
- 3.在数据**稀疏**和**有噪声**的情况下，GLNC 的鲁棒性如何？





• 数据集*3

- TAL2023 (超过550万学习记录)
- EdNet (超过900万学习记录)
- Eedi2020 (超过1000万学习记录)
- 删除答题记录少于30条的学生

Statistics	Datasets		
	TAL2023	EdNet	Eedi2020
# of Students	18,066	78,430	118,971
# of Items	7,652	12,372	27,613
# of KCs	1,175	141	282
# of Interactions	5,549,635	9,058,114	11,001,297
Avg.interactions per student	307	115	92

• 评价指标

- AUC、ACC、RMSE、 r^2 (R-squared, 衡量模型拟合优度的统计量)

$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

- 回归模型的残差平方和 (SSE)、总平方和 (SST) 和回归平方和 (SSR)

• 对比方法

- 认知诊断: IRT (1991)、PMF (2012)、NCD (2020)
- 知识追踪: EKT (2021)、SAKT (2019)、LPKT (2021)



- 学生表现预测
 - GLNC 模型在 EdNet 数据集上的提升最大，AUC、ACC、RMSE 和 r^2 评价指标分别优于先进方法 2.56%、1.11%、1.86% 和 19.87%
 - GLNC 模型，实现最佳性能
 - 捕获学生的全局认知状态和局部动态有效
 - 数据集不同，知识追踪和认知诊断方法表现不同
 - 全局特征和局部特征在不同的数据集中具有不同的重要性

(a) Results of student performance prediction on TAL2023.

Methods	TAL2023			
	AUC	ACC	RMSE	r^2
IRT (Hambleton et al., 1991)	0.796±.002	0.818±.002	0.359±.002	0.188±.002
PMF (Thai-Nghe et al., 2012)	0.790±.002	0.821±.002	0.360±.002	0.181±.003
NCD (Wang et al., 2020)	0.793±.002	0.823±.002	0.358±.001	0.189±.001
EKT (Liu, Huang et al., 2021)	0.791±.002	0.816±.002	0.363±.001	0.196±.004
SAKT (Pandey & Karypis, 2019)	0.787±.002	0.814±.001	0.364±.001	0.191±.004
LPKT (Shen et al., 2021a)	0.792±.002	0.816±.001	0.363±.001	0.196±.004
GLNC	0.808±.002	0.827±.001	0.352±.003	0.218±.001
Gain	1.51%	0.49%	1.68%	11.22%

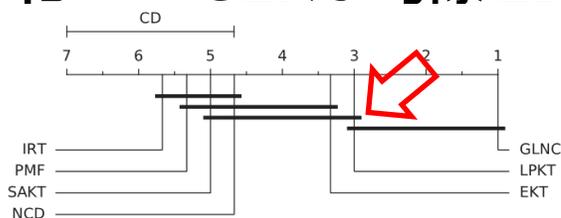
(b) Results of student performance prediction on EdNet.

Methods	EdNet			
	AUC	ACC	RMSE	r^2
IRT (Hambleton et al., 1991)	0.736±.002	0.711±.003	0.433±.002	0.143±.003
PMF (Thai-Nghe et al., 2012)	0.740±.002	0.717±.003	0.434±.002	0.139±.003
NCD (Wang et al., 2020)	0.743±.002	0.723±.003	0.430±.002	0.156±.003
EKT (Liu, Huang et al., 2021)	0.726±.002	0.686±.001	0.451±.001	0.149±.003
SAKT (Pandey & Karypis, 2019)	0.726±.002	0.685±.002	0.452±.001	0.146±.002
LPKT (Shen et al., 2021a)	0.731±.002	0.690±.002	0.449±.001	0.155±.003
GLNC	0.762±.002	0.731±.002	0.422±.001	0.187±.003
Gain	2.56%	1.11%	1.86%	19.87%

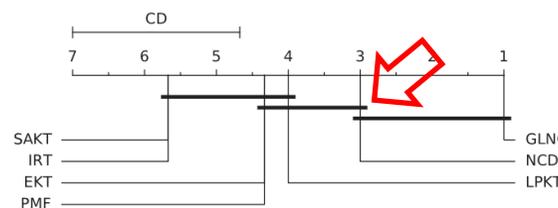
(c) Results of student performance prediction on Eedi2020.

Methods	Eedi2020			
	AUC	ACC	RMSE	r^2
IRT (Hambleton et al., 1991)	0.745±.001	0.703±.002	0.435±.001	0.151±.001
PMF (Thai-Nghe et al., 2012)	0.751±.002	0.714±.001	0.435±.002	0.148±.003
NCD (Wang et al., 2020)	0.749±.001	0.718±.002	0.434±.001	0.158±.005
EKT (Liu, Huang et al., 2021)	0.760±.001	0.715±.001	0.432±.001	0.186±.002
SAKT (Pandey & Karypis, 2019)	0.758±.001	0.714±.001	0.433±.001	0.182±.002
LPKT (Shen et al., 2021a)	0.762±.001	0.716±.001	0.431±.001	0.196±.004
GLNC	0.773±.001	0.730±.001	0.422±.001	0.202±.002
Gain	1.44%	1.63%	2.07%	3.06%

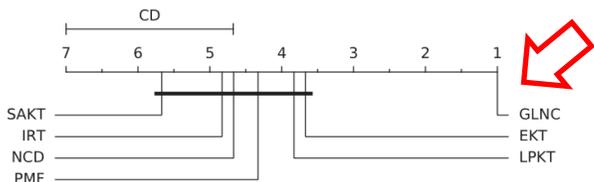
- Nemenyi 结果以临界距离图形式呈现，值表示方法平均排名，排名越低性能越好
 - 交叉线相连意味着该线上的模型之间没有显著差异
 - GLNC 在所有指标上均排名**第一**
 - RMSE：GLNC 优于所有基线
 - ACC：GLNC 优于除 NCD 之外的所有基线
 - AUC 和 r^2 ：GLNC 与除 LPKT 之外的所有基线之间存在明显差异



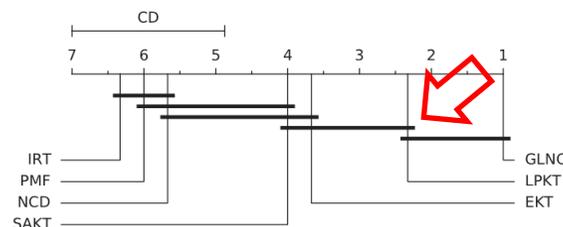
(a) Nemenyi's critical difference diagram on the AUC metric.



(b) Nemenyi's critical difference diagram on the ACC metric.



(c) Nemenyi's critical difference diagram on the RMSE metric.



(d) Nemenyi's critical difference diagram on the r^2 metric.

- 验证GLNC 模型中各个模块的有效性
 - 模块：全局认知建模、局部动态建模、自适应组合和知识概念
 - GLNC w/o g 变体，不测量学生整体认知特征
 - GLNC w/o l 变体，不测量学生局部动态认知特征
 - GLNC w/o c 变体，结合全局和局部认知特征
 - GLNC w/o kc 变体，不考虑练习的知识概念

Table 4
Results of ablation experiments on TAL2023.

Methods	Global feature	Local feature	Adaptive combination	Knowledge concept	AUC	ACC	RMSE	r^2
GLNC w/o g	✗	✓	✓	✓	0.793	0.823	0.358	0.195
GLNC w/o l	✓	✗	✓	✓	0.800	0.824	0.355	0.207
GLNC w/o c	✓	✓	✗	✓	0.806	0.826	0.353	0.215
GLNC w/o KC	✓	✓	✓	✗	0.802	0.825	0.354	0.209
GLNC	✓	✓	✓	✓	0.808	0.827	0.352	0.218

– 实验结论

- 全局认知模型和局部动态学模型都至关重要，缺一不可
- 练习的知识概念信息有益，可捕获更完整的练习特征

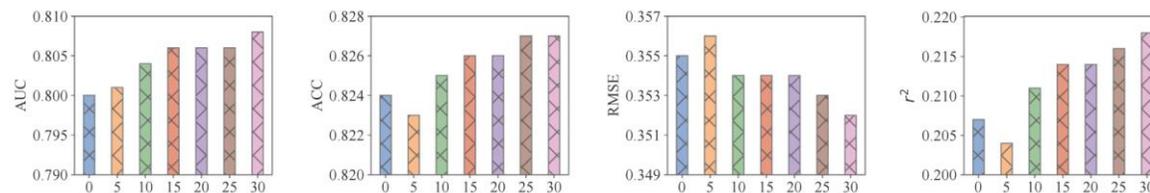
稀疏交互下的鲁棒性验证

– 学生的局部动态建模根据学生最近的**30**次互动完成，**ITS需耗费大量资源**

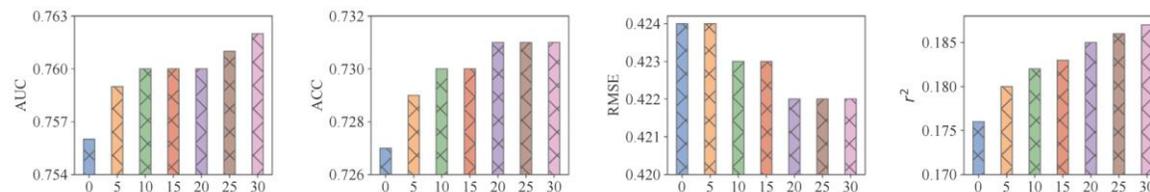
• GLNC 在 7 个不同长度的学生历史交互下表现：**0、5、10、15、20、25 和 30**

– 实验结论

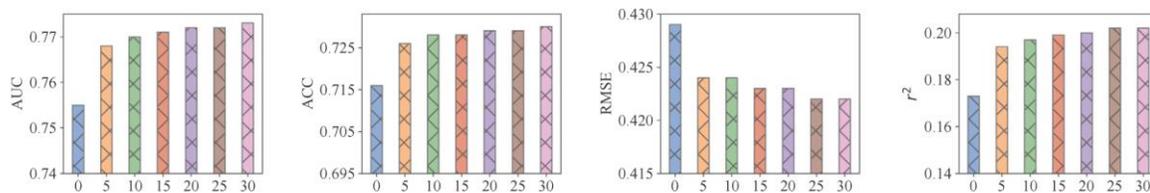
- 学生历史交互越**少**，GLNC 性能越**差**
- GLNC具有良好的稳定性和鲁棒性
 - 仅考虑最近 **5** 次交互，优于最佳基线
- GLNC可以**适应不同ITS**，竞争力强



(a) The effect of the length of students' historical interactions on GLNC's performance for the dataset TAL2023.

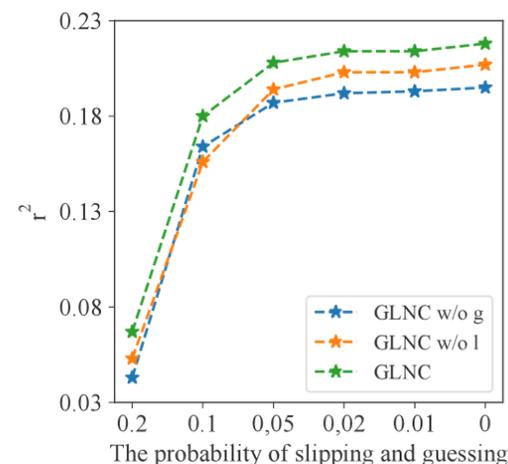
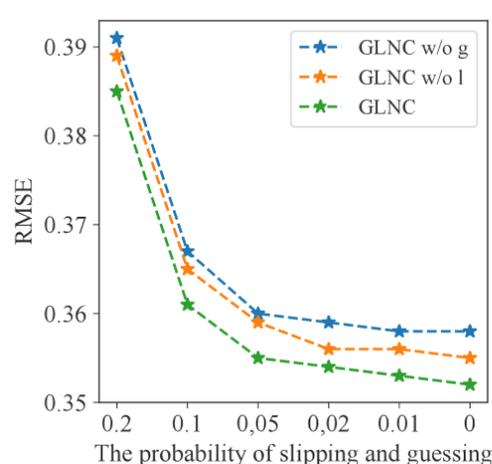
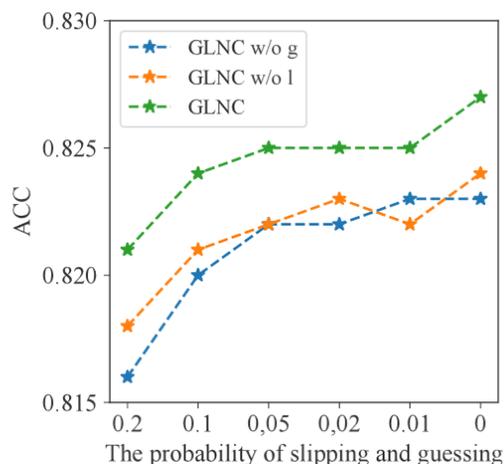
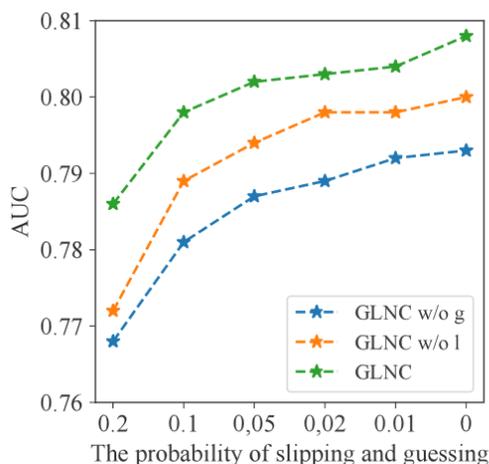


(b) The effect of the length of students' historical interactions on GLNC's performance for the dataset EdNet.



(c) The effect of the length of students' historical interactions on GLNC's performance for the dataset Eedi2020.

- 噪声学生-练习交互数据影响下的鲁棒性
 - 学生-练习交互中的噪音主要来自学生的失误和猜测 (Chen et al., 2022)
 - 训练集中**随机添加**概率分别为 0.2、0.1、0.05、0.02 和 0.01 失误和猜测**噪声**
 - 对比方法: GLNC w/o g (无全局认知) 和 GLNC w/o l (无局部动态认知)
 - 实验结论
 - 添加失误和猜测噪声, 导致 GLNC 及变体性能**下降**
 - 当失误和猜测的概率较高时, GLNC 对噪声数据的影响**更加鲁棒**





特点总结与未来展望

- 算法优势

- ACD

- 心流理论指导，充分挖掘学生认知状态和情感状态间关系
 - 即插即用的情感感知模块，兼容有监督或无监督的学习环境，可集成至 CDM 框架

- GLNC

- 结合学生-练习交互中的全局和局部特征，更全面的认知建模
 - GLNC在稀疏和噪声数据影响下，具有良好的鲁棒性
 - 适应不同ITS，提高学生的学习效率和体验

- 算法劣势

- ACD

- 无监督场景下的**未对**沮丧、困惑、专注、无聊等情感因素具体分析
 - 情感因素**间接**通过参数和交互函数参与认知诊断任务

- GLNC

- 未输出**可解释性**的认知状态（学生对概念的掌握程度）
 - 未考虑学生-问题-概念间复杂关系

- 未来展望

- 深入挖掘无监督场景下的情感因素，使其**数值化**，增强**可解释性**
 - 深度**挖掘**学生-练习-概念间蕴含的复杂**关系**，最大化知识追踪的局部动态认知效果
 - 情感因素**直接参与**认知诊断和知识追踪任务中，提高模型性能
 - 学习路径推荐任务**融入**情感因素和局部-全局学习方式

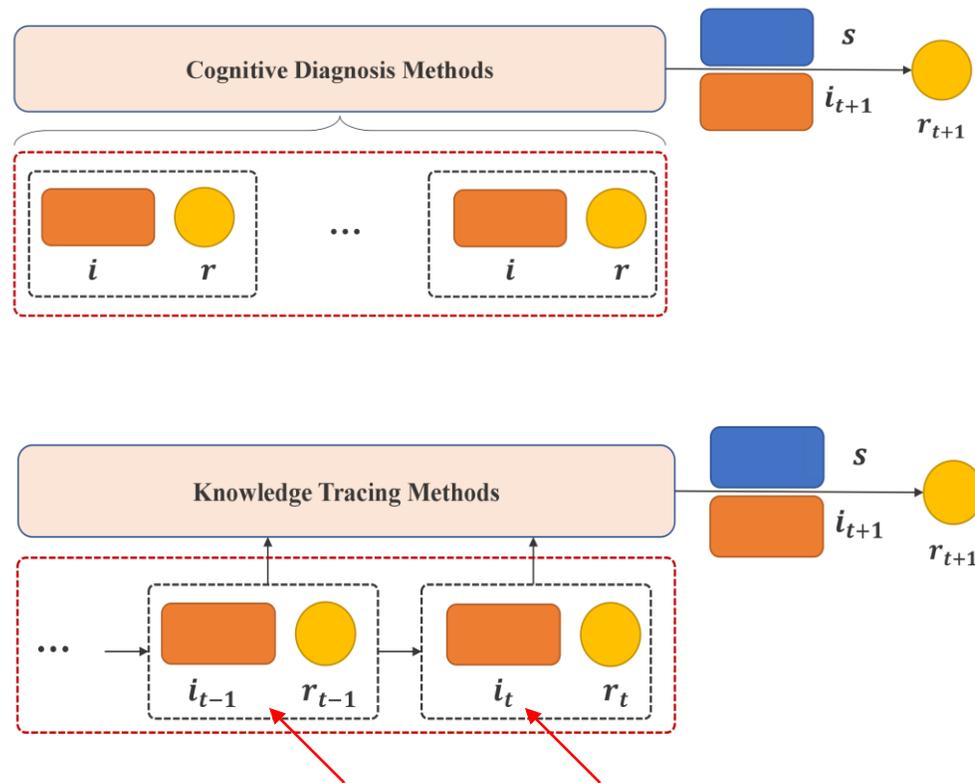
回顾分析

- 预期收获

- 1.了解知识追踪和认知诊断的基本概念和区别
 - 全局静态和局部动态
 - 学生对概念的掌握程度和预测学生正确回答新练习的概率
- 2.熟悉知识追踪和认知诊断的分类及应用场景
 - 心理测量领域和人机交互领域
- 3.理解认知诊断前沿技术的应用原理
 - 考虑新的因素：挖掘学生**认知**状态和**情感**状态间关系
 - 与知识追踪结合：综合考虑**全局静态**特征和**局部动态**特征
- 4.了解认知诊断的发展前景
 - 情感因素**直接参与**认知诊断和知识追踪任务
 - 学习路径推荐任务**融入**情感因素和局部-全局学习方式

• 认知诊断与知识追踪区别

比较维度	认知诊断 (CD)	知识追踪 (KT)
T 目标	评估学生对知识点 (概念) 的掌握程度	预测学生在下一个练习的回答正确率
I 输入	学习记录、Q矩阵	学习记录、Q矩阵
O 输出	学生知识点掌握程度 预测答题情况	预测学生答题情况
模型参数	可解释性	黑盒参数
应用场景	自适应测试 、个性化推荐	学习成绩预测、个性化精准推荐
领域出发点	心理测量	人机交互
特点	全局静态评估 , 结果精细化、多样化	局部动态预测 , 结果时序化



全局静态和局部动态

- [1] Wang, S., Zeng, Z., Yang, X., Xu, K., & Zhang, X. (2024, March). Boosting Neural Cognitive Diagnosis with Student's Affective State Modeling. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 1, pp. 620-627).
- [2] Su, Y., Shen, S., Zhu, L., Wu, L., Huang, Z., Cheng, Z., ... & Wang, S. (2024). Global and local neural cognitive modeling for student performance prediction. *Expert Systems with Applications*, 237, 121637.

知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者，寿。

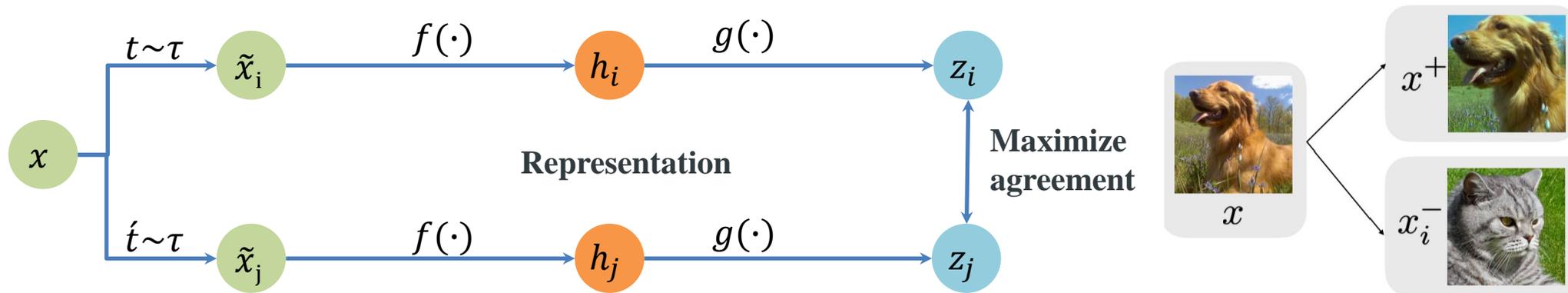
谢谢！



对比学习-自监督学习方法

– 核心思想：代理任务+目标函数

– 代理任务作用阶段 → 特征提取编码器 → MLP层 → 目标函数作用阶段



– 设计模型结构和对比损失，使正样例的表示在表示空间更接近，负样例更远

• 图像：空间/几何扰动和外观/色彩扰动

• NLP：词级别（token-level）{句子剪裁（crop）、删除词/词块（span）、换序、同义词替换} 和表示级别（embedding-level）{高斯噪声、dropout}