

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



认知诊断技术及其研究

杨晓楠

2023年09月03日

- **总结反思**

- 演讲流畅度不高，语气稍显平淡
- 基础概念讲解内容过细

- **相关内容**

- 周瑾洁《从赋能学习到知识追踪》——2022.09.18
- 周瑾洁《基于图的知识追踪方法研究》——2023.05.07

- 背景简介
- 基础概念
 - 认知诊断
 - 基本模型
- 算法原理
 - NeuralCDM
 - RCD
- 应用总结
- 参考文献

- 预期收获
 - 了解认知诊断的定义
 - 了解智能教育领域下的认知诊断研究脉络
 - 理解基于神经网络的认知诊断的方法原理
 - 了解认知诊断的前沿发展

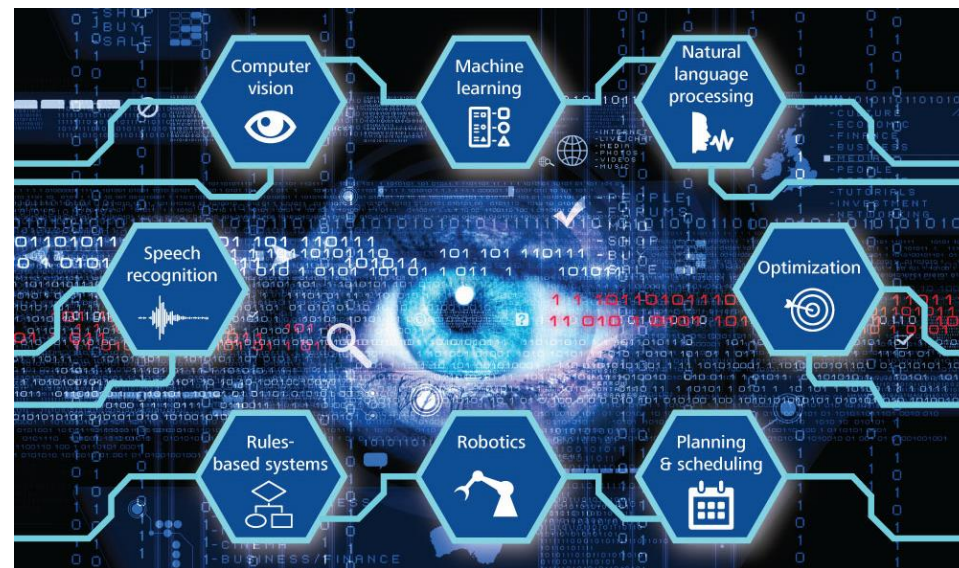
- 认知

- 一种意识活动，是个体认识客观世界的**信息加工过程和结果**，是对事物概念的判断和规律的总结
- 包括感知、记忆、联想（记忆激活）、思维、想象，推理，决策等



- 认知智能

- 人工智能发展的三个阶段：运算智能，感知智能和认知智能
- 出发点：人工智能能否像**人类一样思考**？
- 人工智能正从**感知智能向认知智能**演进
- 两大核心技术：知识图谱技术和自然语言处理技术



• 智能教育

- 人工智能在教育领域蓬勃发展
- 目标：智能发现教学盲点并扫除，提供**即时、全面针对性**教学，促进学生全面发展
- 应用：

- 学业状态诊断
- 学业水平预测
- 自适应学习路径推荐
- 学业预警
- 学业干预

感知智能



- **学生认知状态诊断**
- 认知水平预测
- 自适应认知重塑路径推荐
- 认知预警
- 认知干预

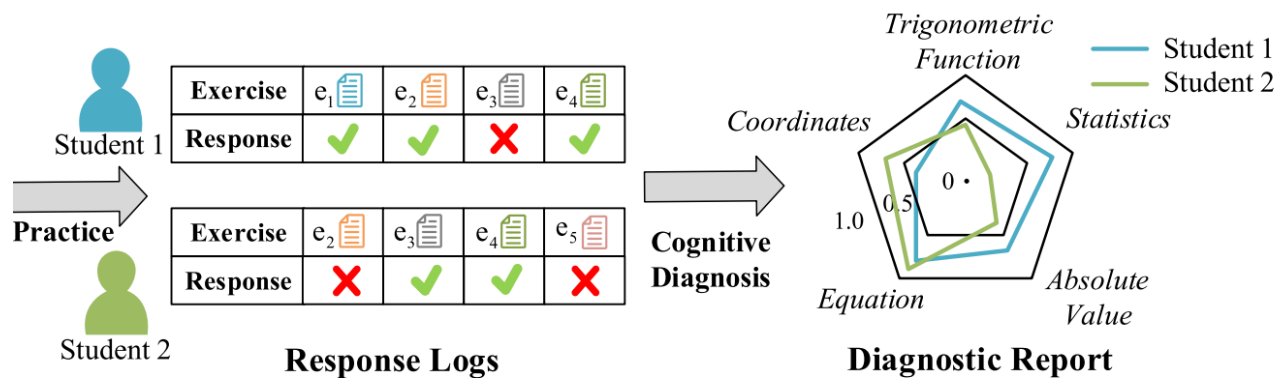
认知智能

- 认知诊断 (Cognitive Diagnosis, CD)

- 定义：通过观察个体的行为，衡量被测个体的心理或活动行为规律，对被测个体认知的优缺点进行综合评估和诊断，进而有针对性的促进被测个体的全面发展
- 应用领域：医疗诊断、教育教学等

- 智能教育领域下的认知诊断

- 对学习者的某一给定时间段的学习行为进行研究分析，分析挖掘基于这些行为的认知状态(如对所需技能的熟练程度)



- ① 学生1、2分别做了1-5练习题 (分值20*5)，最终得分相同
- ② 学生1、2认知状态是否相同?
- ③ 不同练习之间的区别与联系?

• 问题定义

- 学生集合 S : $S = \{s_1, s_2, s_3, \dots, s_N\}$
- 试题练习集合 E : $E = \{e_1, e_2, e_3, \dots, e_M\}$
- 知识点集合 K : $K = \{k_1, k_2, k_3, \dots, k_K\}$
- 答题记录

k_1 : <i>Trigonometric Function</i>	k_2	k_3	k_4	k_5	
k_2 : <i>Statistics</i>					}
k_3 : <i>Coordinates</i>					
k_4 : <i>Equation</i>					
k_5 : <i>Absolute Value</i>					

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{matrix}$$

$$R = \{(s, e, r) | s \in S, e \in E\}$$

- Q矩阵

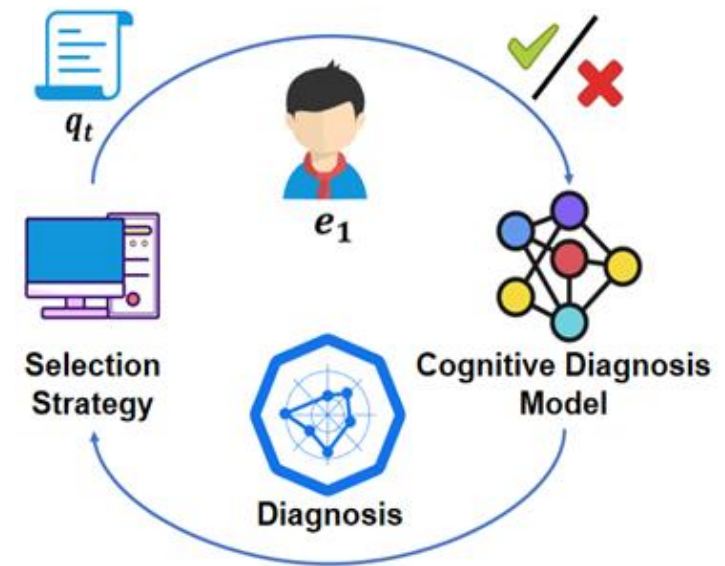
- 反映练习与知识点间的映射关系

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

- 问题：给定学生答题记录和Q矩阵，通过学生答题预测的过程获取学生在各个知识点上的掌握程度

- 能力水平研究范式

- 强调对被试宏观层次的**能力水平**测量和评估
- 经典测量理论(Classical Test Theory, CTT)
 - 直接使用被试者对项目的**应答得分**衡量被试的认知状态



- 认知水平研究范式

- 强调理解个体**内部微观层次**的**认知加工**过程
- 项目反应理论 (Item Response Theory, IRT)
 - 模型交互函数

$$P(X_{ij} = 1 | \theta_i, \alpha_j, \beta_j, c_j) = c_j + (1 - c_j) \frac{e^{D\alpha_j(\theta_i - \beta_j)}}{1 + e^{D\alpha_j(\theta_i - \beta_j)}}$$

学生能力

练习区分度

练习难度

猜测度

- 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 在试题 j 的作答情况

$$n_{ij} = \prod_{k=1}^K \alpha_{ik}^{Q_{jk}}$$

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

- 概率预测函数

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

s_j : 学生在掌握了试题 j 所考察的所有知识点的情况下做错的概率

g_j : 学生在并不完全掌握试题 j 所考察的所有知识点下猜对的概率



NeuralCDM

T	目标	诊断学生对特定 知识概念 的熟练程度
I	输入	学生练习和考试日志 (学生记录*16k、练习文本*930k、知识概念*1.5k)
P	处理	<ol style="list-style-type: none">1. 计算学生知识点熟练度向量2. 计算知识点相关度向量、知识点难度向量和练习区分度向量3. 通过神经网络模拟学生与练习之间的交互建模4. 通过交叉熵损失函数训练神经网络框架5. 预测学生试题回答结果
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	现有方法依赖于人工设计的交互函数，不足以捕捉学生和练习之间的复杂关系
C	条件	单调性假设，基于专家知识的Q矩阵
D	难点	如何通过神经网络建模学生和练习之间的交互，同时保留其可解释性
L	水平	AAAI 2020 (CCF-A)



- 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

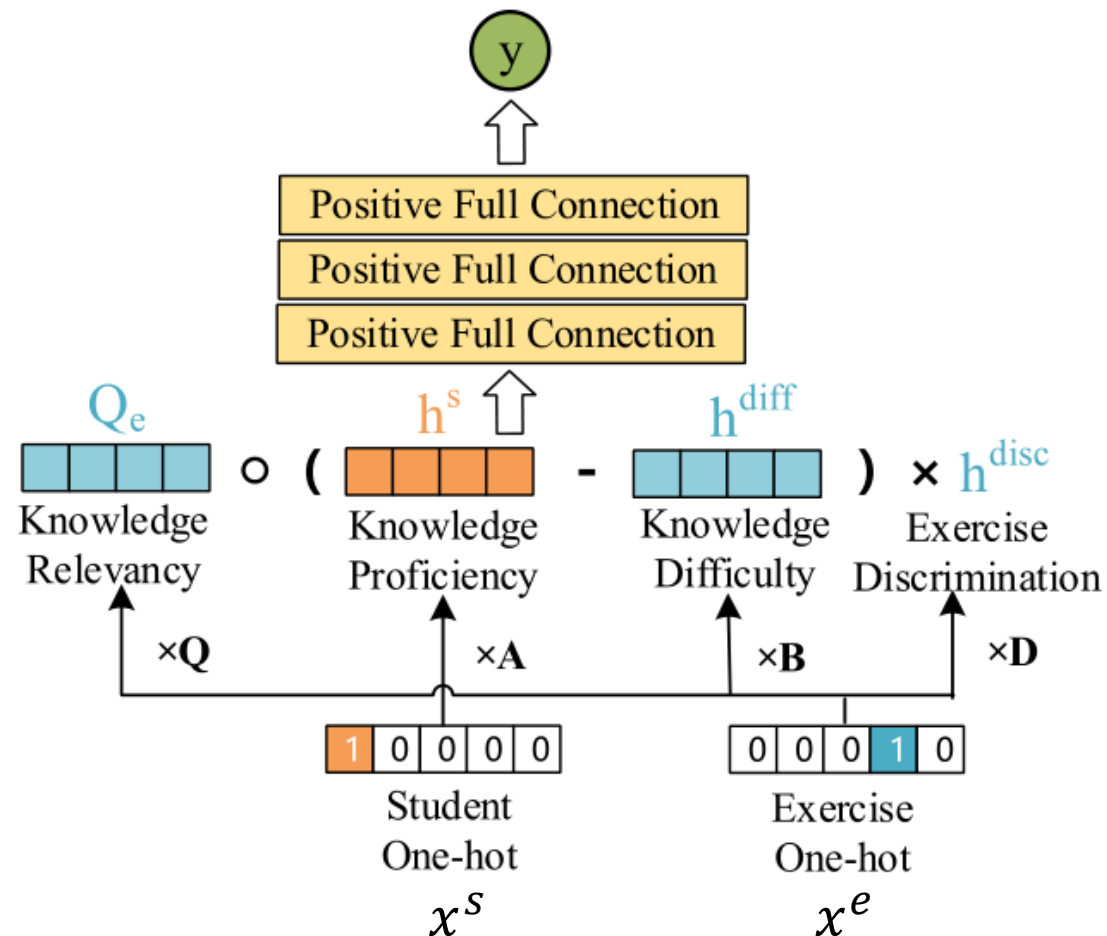
$$h^{diff} = \text{sigmoid}(x^e \times B)$$

- 练习区分度向量

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

- 交互层输入

$$x = Q_e \circ (h^s - h^{diff}) \times h^{disc}$$



- 学生-练习交互建模

- 单调性假设

- 学生答对练习的概率随学生任意一维知识点熟练度单调递增

- 交互函数

$$f_1 = \text{sigmoid}(W_1 \times x^T + b_1)$$

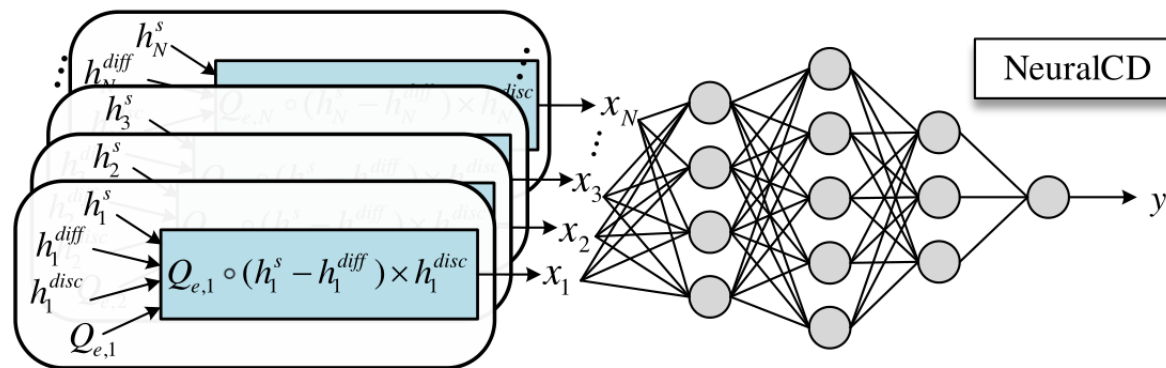
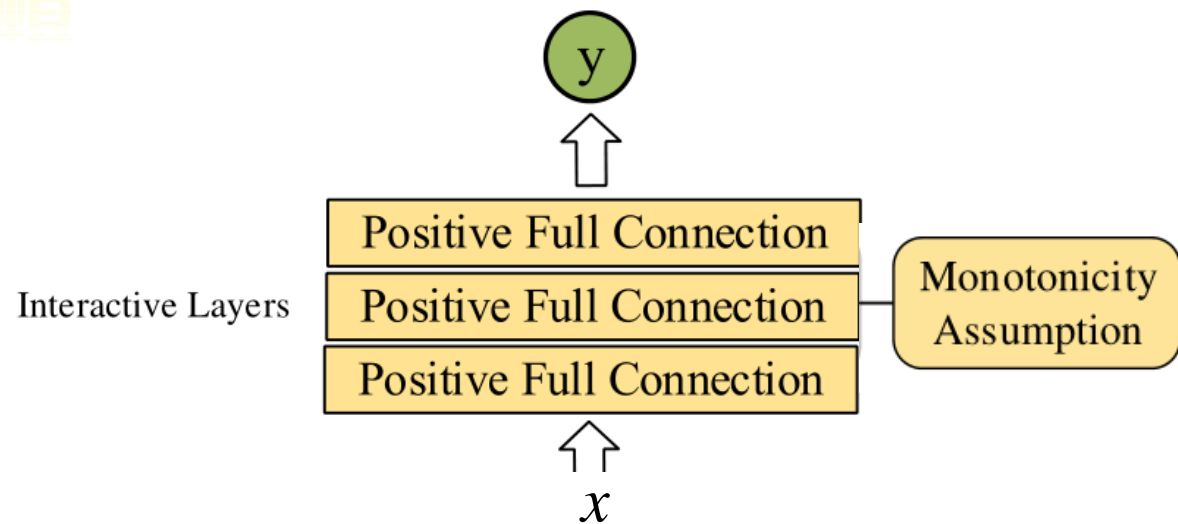
$$f_2 = \text{sigmoid}(W_2 \times f_1 + b_2)$$

$$y = \text{sigmoid}(W_3 \times f_2 + b_3)$$

其中 $W_1, W_2, W_3 \geq 0, \frac{\partial y}{\partial h_i^s} \geq 0$

- 损失函数

$$\text{loss}_{CDM} = - \sum_i (r_i \log y_i + (1 - r_i) \log(1 - y_i))$$



- 利用试题文本对NeuralCDM扩展 (NeuralCDM+)

- 问题：手工标记的Q矩阵可能存在不可避免的**误差和主观偏差**

- 解决方案：

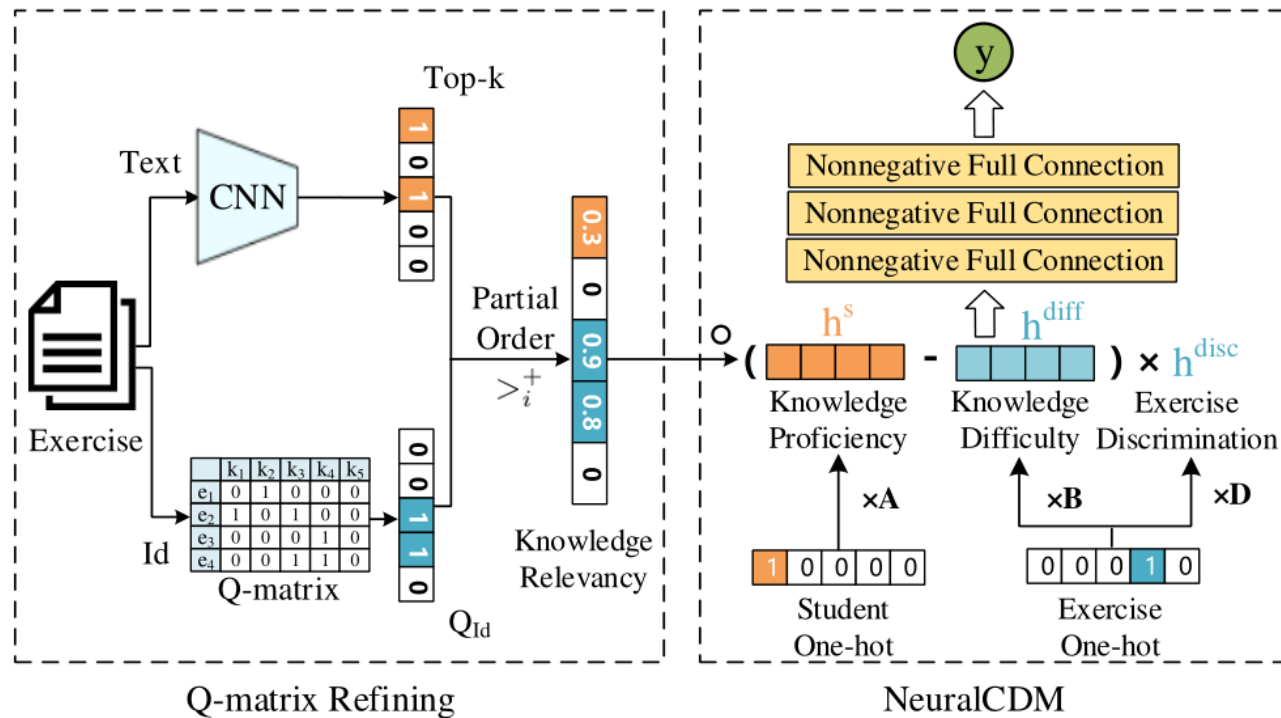
- 利用**练习文本内容**挖掘练习中被忽略的知识概念来细化Q矩阵

- 操作

- 使用CNN预训练模型预测与输入练习相关的**top-k**个知识概念
- 结合专家标注的Q矩阵进行优化

$$a >_i^+ b, \text{ if } Q_{ia} = 1 \text{ and } Q_{ib} = 0 \text{ and } b \in V_i^k$$

Q矩阵标注 ≥ 预测top-k ≥ 其他=0



实验设计

– 数据集

- Math
- ASSIST (未提供练习文本)

– 评价指标

- 预测: Accuracy、RMSE、AUC
- 可解释性: 一致性度量

Dataset	Math	ASSIST
#Students	10,268	4,163
#Exercises	917,495	17,746
#Knowledge concepts	1,488	123
#Response logs	864,722	324,572
#Knowledge concepts per exercise	1.53	1.19

$$DOA(k) = \frac{1}{Z_1} \sum_{a=1}^N \sum_{b=1}^N I(F_{ak}^s > F_{bk}^s) \frac{\sum_{j=1}^M I(Q_{jk} = 1) \wedge J(j, a, b) \wedge I(r_{aj} > r_{bj})}{Z_0}$$

其中

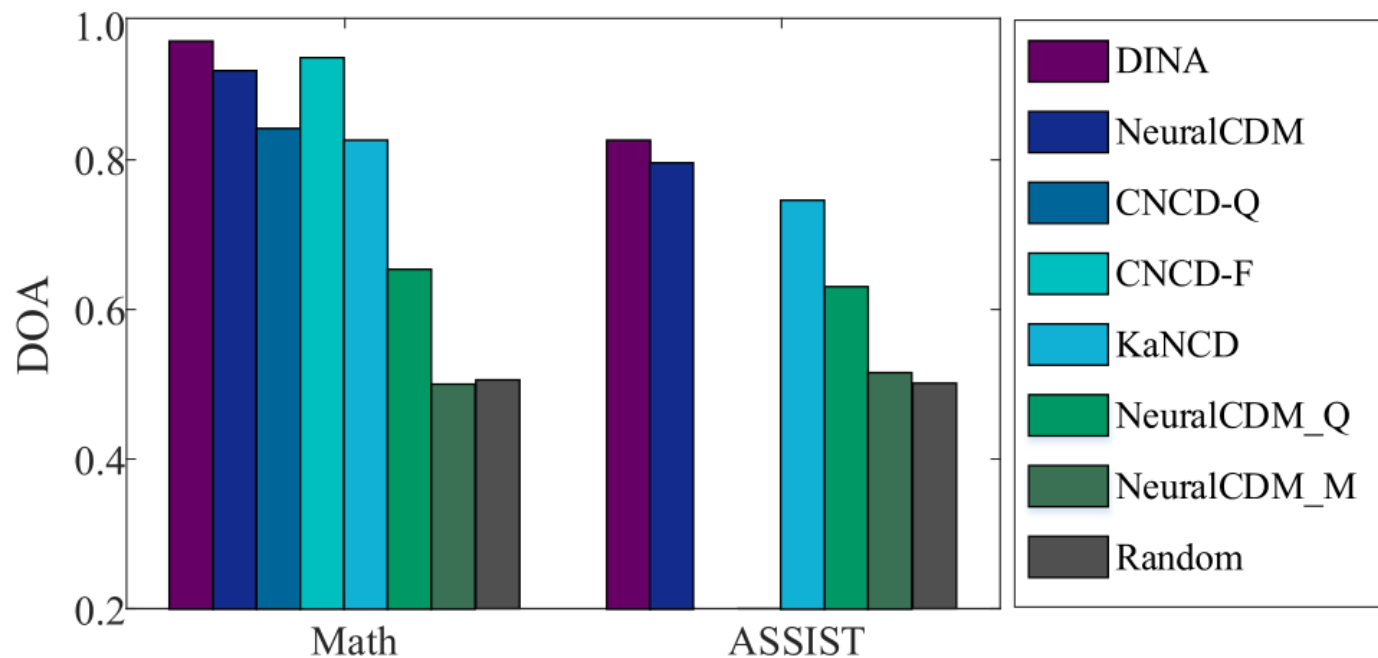
$$Z_0 = \sum_{j=1}^M I(Q_{jk} = 1) \wedge J(j, a, b) \wedge I(r_{aj} \neq r_{bj}); Z_1 = \sum_{a=1}^N \sum_{b=1}^N I(F_{ak}^s > F_{bk}^s) I(Z_0 > 0)$$

- 学生成绩预测

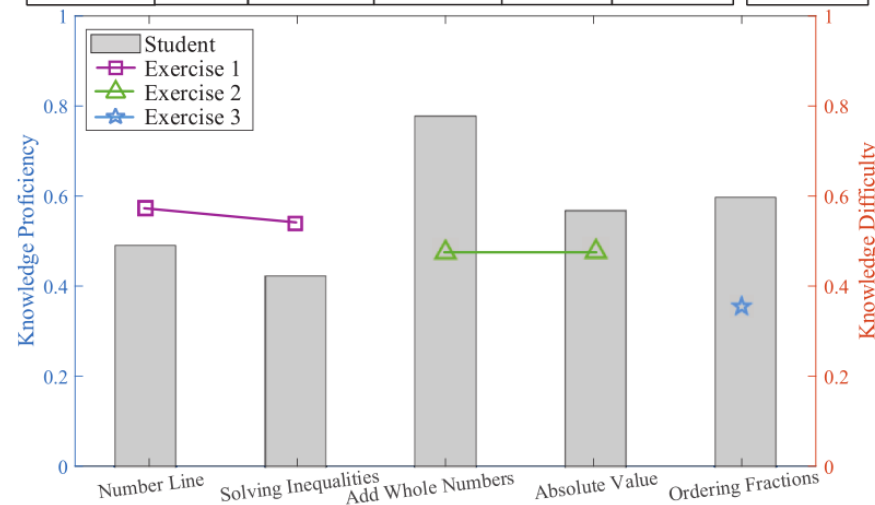
- 对比传统方法：DINA、IRT、MIRT、PMF
- 在所有指标上，NeuralCDM/CDM+性能均优于其他四种技术
- NeuralCDM+优于NeuralCDM，证明了Q矩阵优化方法的有效性

Model	Math			ASSIST		
	Accuracy	RMSE	AUC	Accuracy	RMSE	AUC
DINA	0.593±.001	0.487±.001	0.686±.001	0.650±.001	0.467±.001	0.676±.002
IRT	0.782±.002	0.387±.001	0.795±.001	0.674±.002	0.464±.002	0.685±.001
MIRT	0.793±.001	0.378±.002	0.813±.002	0.701±.002	0.461±.001	0.719±.001
PMF	0.763±.001	0.407±.001	0.792±.002	0.661±.002	0.476±.001	0.732±.001
NeuralCDM	0.792±.002	0.378±.001	0.820±.001	0.719±.008	0.439±.002	0.749±.001
NeuralCDM+	0.804±.001	0.371±.002	0.835±.002	-	-	-

- 诊断结果的**可解释性**
 - NeuralCDM/CDM+高于所有基线，证明诊断的知识水平是合理的
- 案例分析
 - 当学生的熟练程度满足练习的要求时，更有可能做出正确的反应



	Number Line	Solving Inequalities	Add Whole Numbers	Absolute Value	Ordering Fractions	Student Response
Exercise 1	1	1	0	0	0	✗
Exercise 2	0	0	1	1	0	✓
Exercise 3	0	0	0	0	1	✓



• 优势

- 首次提出一个神经认知诊断框架，自动学习复杂交互作用，并保留其可解释性
- 考虑练习文本中的信息，扩展框架提高诊断效果

• 劣势

- 知识概念预测模型有效性不足
- 对神经网络权值的正向限制可能会限制其近似能力
- 未考虑概念之间的相关性，如先决条件等





RCD

T	目标	诊断学生对特定 知识概念 的熟练程度
I	输入	学生练习和考试日志 (学生记录*12k、练习文本*25k、知识概念*1k, 概念依赖关系*4k)
P	处理	<ol style="list-style-type: none"> 1. 学生-练习-概念多层关系图构建 2. 概念、练习和学生的矢量化嵌入表示 3. 通过多层神经网络对不同层嵌入进行融合 4. 通过诊断网络预测学生练习的反应分数, 推断学生的认知水平
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	现有方法未考虑概念间的相互依赖关系, 缺乏对学生-练习-概念的层次关系全面建模
C	条件	基于专家知识的概念依赖关系
D	难点	如何融合来自不同层次关系图的信息, 有效地学习关系感知表示
L	水平	ACM SIGIR 2021 (CCF-A)

- 概念依赖关系图

- 依赖关系: **先决条件、相似性**

- 关系图 $G_C(C, R_C)$, 其中关系集合

$$R_C = \{R_C^r, r = 1, \dots, R\}$$

- 练习-概念关联图

- 关联图 $G_{qc}(Q \cup C, R_{qc})$, 其中练习-概念关联集合

$$R_{qc} = \{r_{q_i \leftrightarrow c_j} | q_i \in Q, c_j \in C\}$$

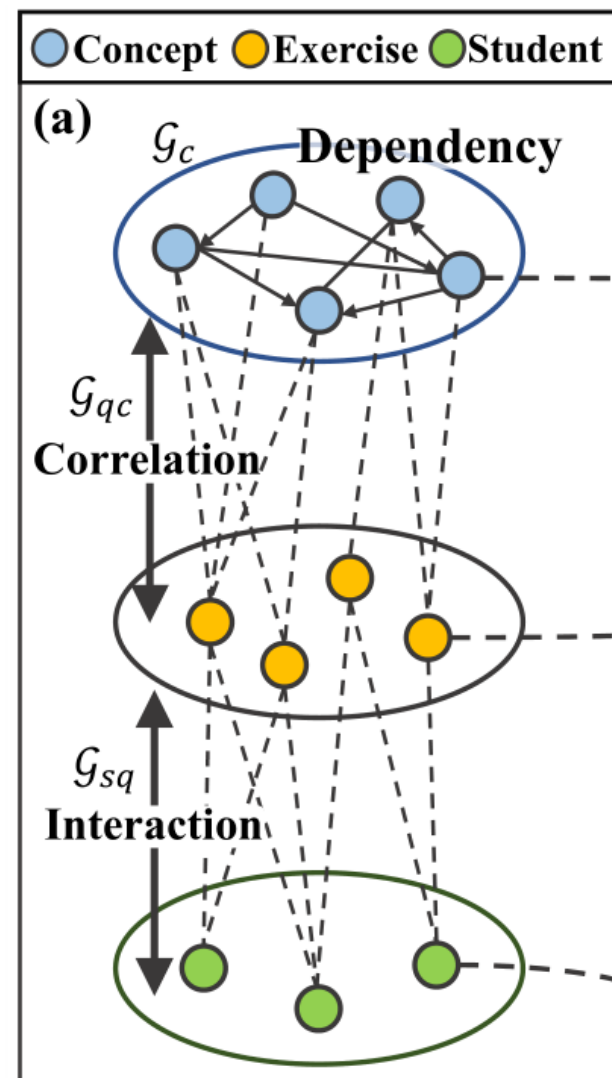
- 学生-练习交互图

- 交互图 $G_{sq}(S \cup Q, R_{sq})$, 其中学生-练习交互集合

$$R_{sq} = \{r_{s_i \leftrightarrow q_j} | s_i \in S, q_j \in Q\}$$

- **学生-练习-概念层次关系图**

- 关系图 $G(G_C \cup G_{qc} \cup G_{sq}, R_C \cup R_{qc} \cup R_{sq})$



- 嵌入表示

- 对学生、练习和知识概念进行**编码**

$$s_z^0 = x_z^T S, q_d^0 = x_d^T Q, c_k^0 = x_k^T C$$

- 知识概念融合

- 融合来自**概念间**和**练习**的相关信息

相关强度:

$$c_k^{l+1} = c_k^l + \sum_{r \in R_c} \eta_{kr}^{l+1} \tilde{p}_r^{l+1} + \eta_{k1}^{l+1} \tilde{e}_k^{l+1}$$

$$\eta_{kr}^{l+1} = F_{cp}([c_k^l, \tilde{p}_r^{l+1}])$$

$$\eta_{k1}^{l+1} = F_{ce}([c_k^l, \tilde{e}_k^{l+1}])$$

概念相关:

$$\tilde{p}_r^{l+1} = \sum_{a \in N_k^r} \alpha_{ka}^{l+1} W_c^l c_a^l$$

$$\alpha_{ka}^{l+1} = F_{cc}([W_c^l c_k^l, W_c^l c_a^l])$$

练习相关:

$$\tilde{e}_r^{l+1} = \sum_{b \in N_k^q} \beta_{kb}^{l+1} W_q^l q_b^l$$

$$\beta_{kb}^{l+1} = F_{cq}([W_q^l c_k^l, W_q^l q_b^l])$$

练习融合

– 融合来自概念和学生的相关信息

$$q_d^{l+1} = q_d^l + \gamma_{d1}^{l+1} \tilde{u}_a^{l+1} + \gamma_{d2}^{l+1} \tilde{v}_b^{l+1}$$

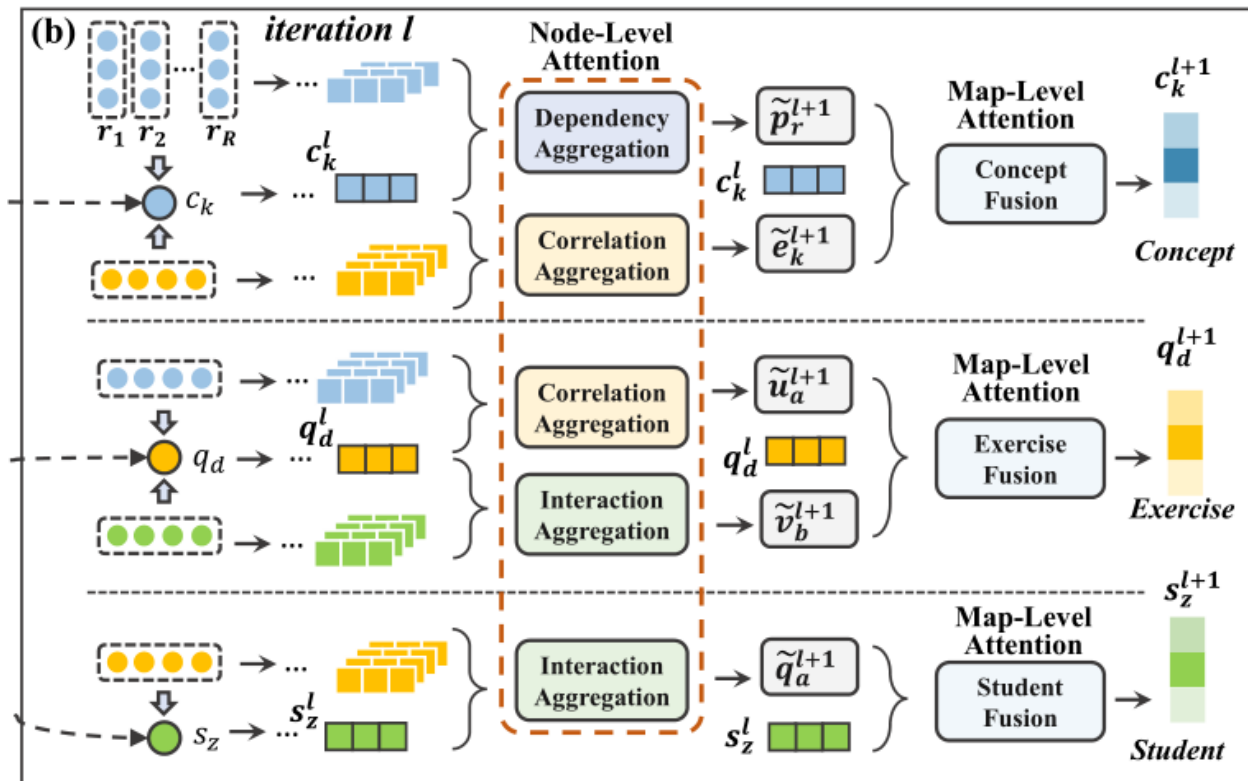
$$\tilde{u}_d^{l+1} = \sum_{a \in N_d^c} \mu_{da}^{l+1} W_{cu}^l c_a^l$$

$$v_d^{l+1} = \sum_{b \in N_d^s} \nu_{db}^{l+1} W_s^l s_b^l$$

学生融合

– 融合来自练习的相关信息

$$s_z^{l+1} = s_z^l + \tilde{q}_a^{l+1}; \quad \tilde{q}_a^{l+1} = \sum_{a \in N_z^q} \rho_{za}^{l+1} W_{qs}^l q_a^l$$



- 认知水平诊断

- 学生因素（如能力水平）

$$h_z^{(k)} = \sigma(f_{sc}(s_z \oplus c_k))$$

- 练习因素（如难度水平）

$$h_d^{(k)} = \sigma(f_{qc}(q_d \oplus c_k))$$

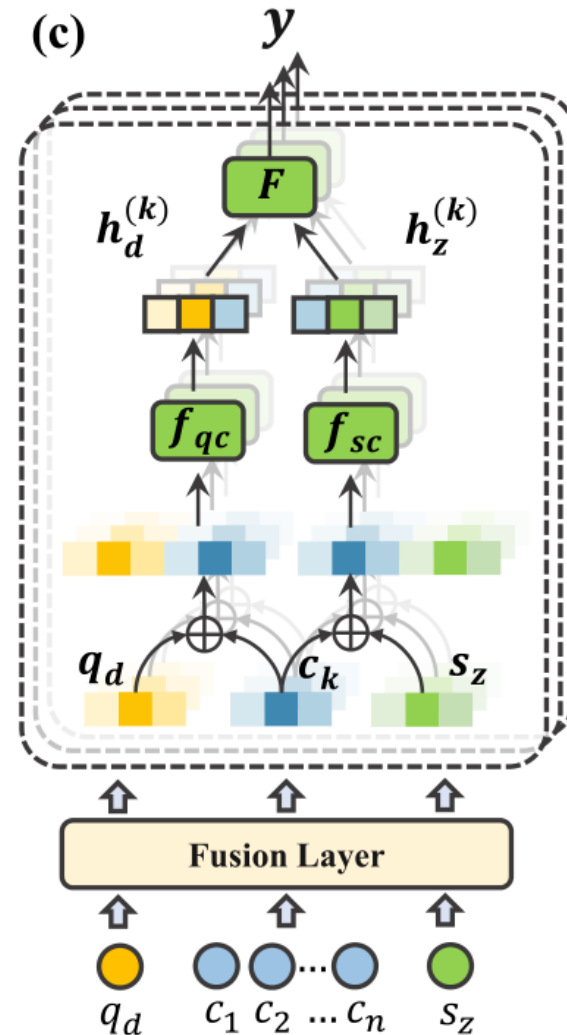
- 量化学生-练习之间的优势效用

- 归一化为概率预测并有选择的累积

$$y_{q_d} = \frac{1}{n} \sum_{c_k} \sigma(F(h_z^{(k)} - h_d^{(k)}))$$

- 损失函数

$$loss_{RCD} = - \sum_i (r_i \log y_i + (1 - r_i) \log(1 - y_i))$$



• 问题

- 1. 与现有的 CD 模型相比，我们提出的 RCD 在**准确性**方面表现如何？
- 2. 当将 RCD 混合交互和结构信息用于当前方法时，模型的性能如何？
- 3. RCD 对多层次学生-练习-概念交互建模的有效性如何？
- 4. RCD 在关系图中捕捉学生-练习-概念交互的能力如何？
- 5. RCD 对诊断学生知识状态的**解释**如何？



实验设计

– 数据集

- Junyi (包含概念依赖关系)
- ASSIST

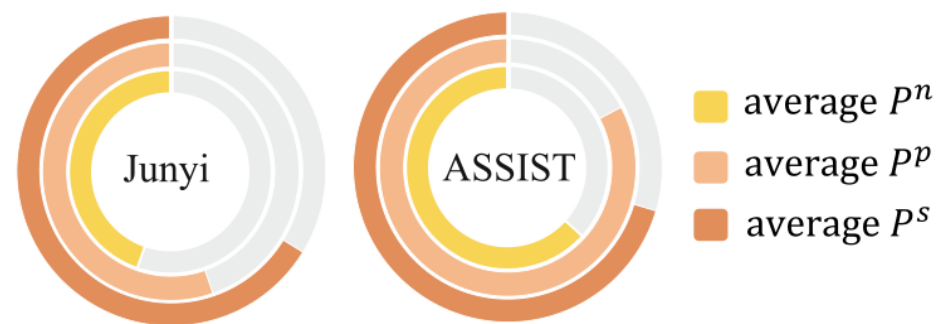
– 评价指标

- Accuracy、RMSE、AUC

概念依赖关系的影响

- 无依赖关系的回答正确性概率 P^n
- 具有先决关系的回答正确性概率 P^p
- 具有相似关系的回答正确性概率 P^s

Statistics	Junyi	ASSIST
#Students	10,000	2,493
#Exercises	835	17,746
#Knowledge concepts	835	123
#Response records	353,835	267,415
#Knowledge concepts per exercise	1	1.20
#Response records per student	35.38	107.27
#Prerequisite relations among concepts	988	1,164
#Similarity relations among concepts	1,040	1,256





• 学生成绩预测

- 对比方法：IRT、MIRT、MF、NeuralCDM (NCD)
- 将结构信息融合到对比方法中 (R-xxx)
- RCD在所有训练测试比例下都表现最好 (Q1)
- 几乎所有基于关系图的方法都优于对应基线的表现 (Q2)

(a) Junyi

Train/Test ratio		50%/50%			60%/40%			70%/30%			80%/20%		
Methods		ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE
Baseline	IRT	67.56%	75.12%	44.53%	67.46%	76.09%	43.65%	67.45%	76.66%	43.13%	67.60%	77.50%	42.68%
	MIRT	74.70%	79.16%	41.43%	74.83%	79.32%	41.36%	74.81%	79.47%	41.35%	75.13%	79.89%	41.17%
	MF	66.87%	74.43%	45.20%	67.35%	74.98%	44.79%	67.81%	75.46%	44.37%	68.34%	76.44%	43.73%
	NCD	74.21%	78.68%	42.01%	74.40%	79.05%	41.67%	74.30%	78.95%	41.78%	74.43%	79.09%	41.72%
Relation-based	R-IRT	74.64%	79.12%	41.51%	74.67%	79.18%	41.44%	74.71%	79.35%	41.39%	74.68%	79.43%	41.36%
	R-MIRT	76.60%	81.82%	40.14%	76.59%	81.71%	40.15%	76.67%	81.87%	40.15%	76.48%	82.23%	40.11%
	R-MF	74.38%	78.79%	41.63%	74.33%	78.75%	41.66%	74.54%	78.85%	41.60%	74.43%	79.06%	41.59%
	R-NCD	74.48%	78.94%	41.90%	74.56%	78.86%	42.00%	74.61%	79.07%	41.85%	74.66%	79.31%	41.59%
	RCD	76.61%	81.84%	40.05%	76.66%	81.88%	40.10%	77.01%	82.26%	39.82%	77.16%	82.62%	39.63%

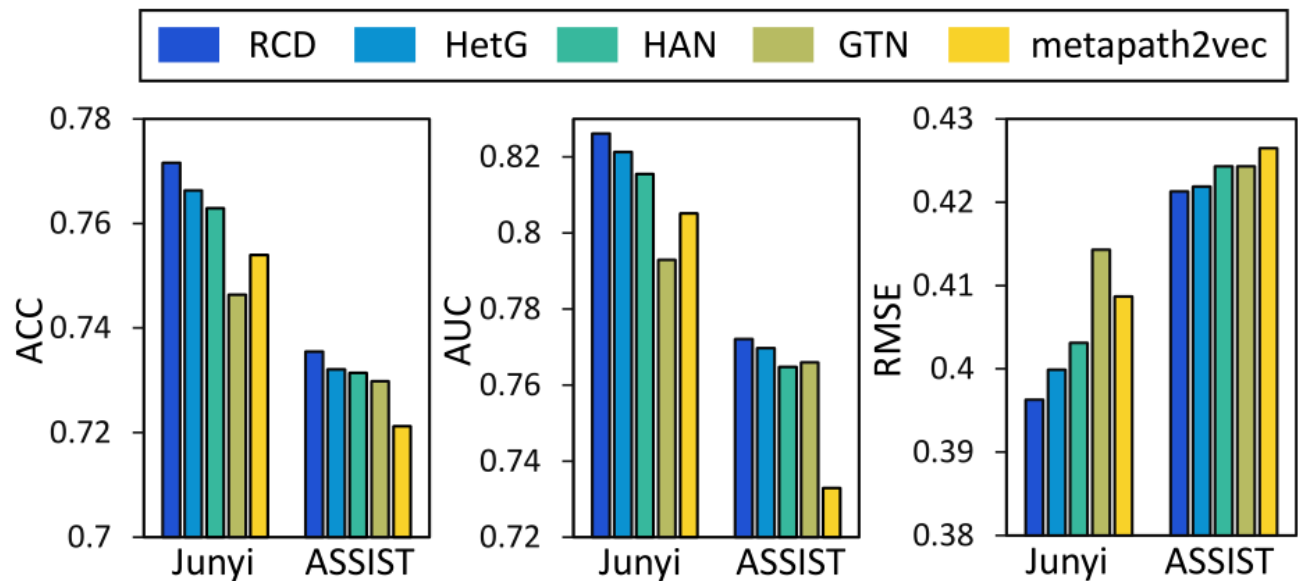
(b) ASSIST

Train/Test ratio		50%/50%			60%/40%			70%/30%			80%/20%		
Methods		ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE
Baseline	IRT	63.46%	66.46%	49.10%	63.76%	67.81%	48.13%	64.00%	68.79%	47.36%	64.26%	69.83%	46.59%
	MIRT	67.41%	67.85%	47.98%	68.53%	71.18%	48.39%	70.39%	72.83%	46.03%	71.70%	74.94%	45.17%
	MF	61.84%	70.09%	49.34%	64.23%	72.85%	47.13%	65.67%	74.62%	45.86%	67.12%	76.45%	44.51%
	NCD	71.51%	73.77%	43.90%	71.85%	74.21%	43.81%	72.17%	74.92%	43.78%	73.14%	75.94%	43.08%
Relation-based	R-IRT	69.53%	69.43%	45.15%	70.44%	72.59%	44.11%	71.43%	74.08%	43.58%	69.53%	69.43%	45.15%
	R-MIRT	72.82%	76.13%	42.59	72.50%	76.18%	42.73%	72.90%	76.55%	42.49%	73.40%	77.20%	42.22%
	R-MF	66.99%	66.90%	46.23%	67.54%	67.53%	45.98%	71.96%	74.85%	44.45%	71.61%	74.35%	43.70%
	R-NCD	71.46%	74.42%	43.61%	71.24%	73.84%	43.73%	71.78%	74.62%	43.75%	72.39%	75.32%	43.25%
	RCD	72.87%	76.24%	42.50%	72.82%	76.19%	42.54%	73.11%	76.63%	42.41%	73.55%	77.21%	42.13%

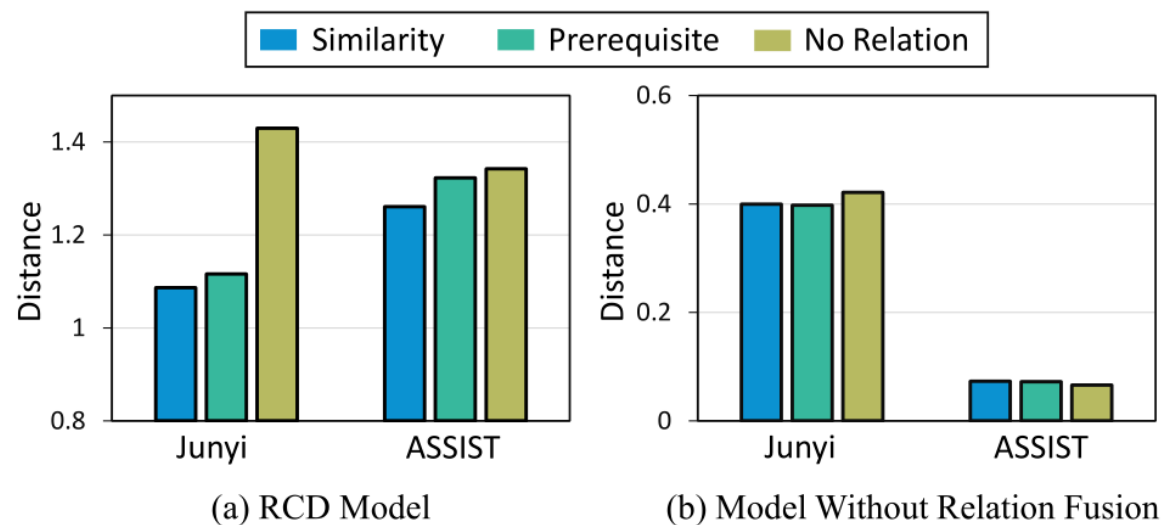
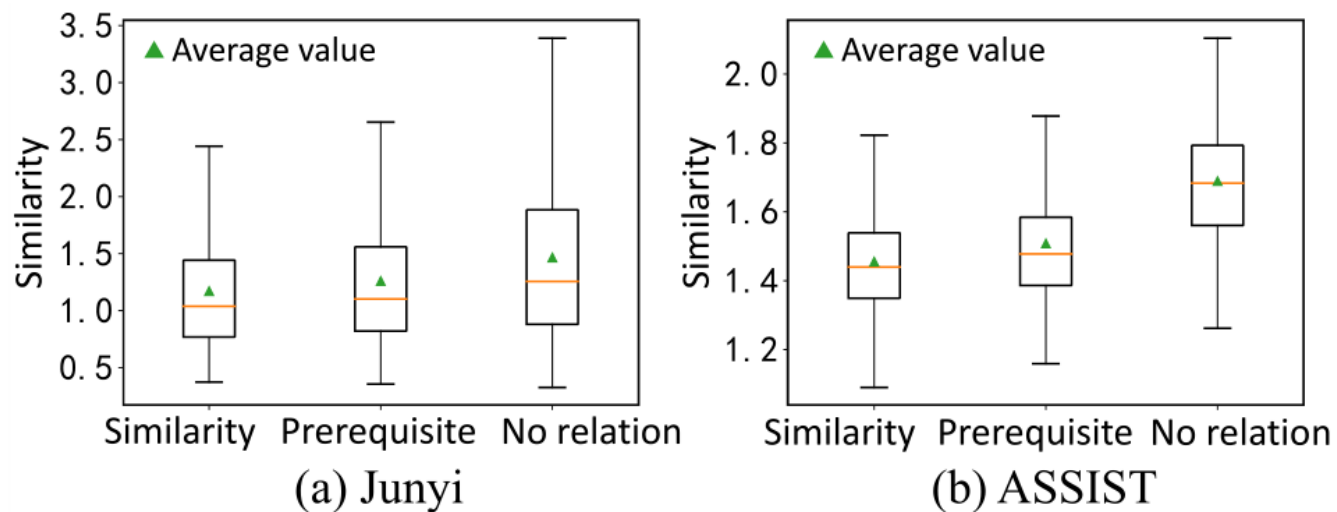
- 多层次关系建模的有效性 (Q3)

- RCD与几种有竞争力的异质图表示学习方法在成绩预测任务上进行比较
- 在融合层采用**不同的图学习方法** (即HetG、HAN、GTN和metapath2vec), 并设置**相同的诊断层**

- RCD分数最高, 表明多层次的融合建模适合于建立多层学生-练习-概念的互动模型
- HAN和HetG也取得了良好的性能, 表明RCD诊断模型具有突出的建模能力



- 概念依赖关系捕捉 (Q4)
 - 使用欧氏距离衡量每对概念嵌入之间的相似性
 - RCD可以很好地捕捉概念之间的关系
- 练习-概念相关性捕捉 (Q4)
 - 计算每对习题嵌入的欧氏距离
 - RCD关系融合模型, 相似的、先决条件的和没有关系的习题之间的距离呈现明显的增加趋势
 - RCD可以很好地学习习题和知识概念之间的关联



• 学生-练习交互捕捉 (Q4)

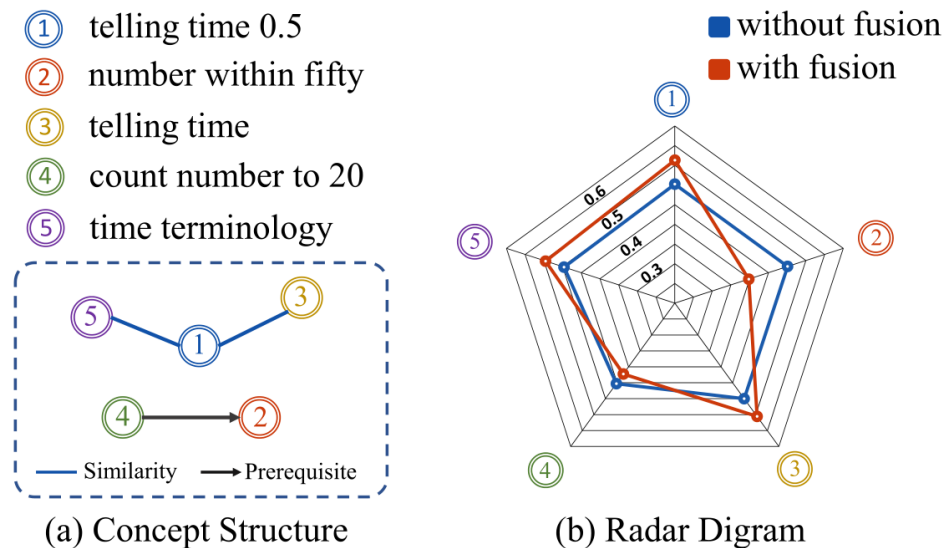
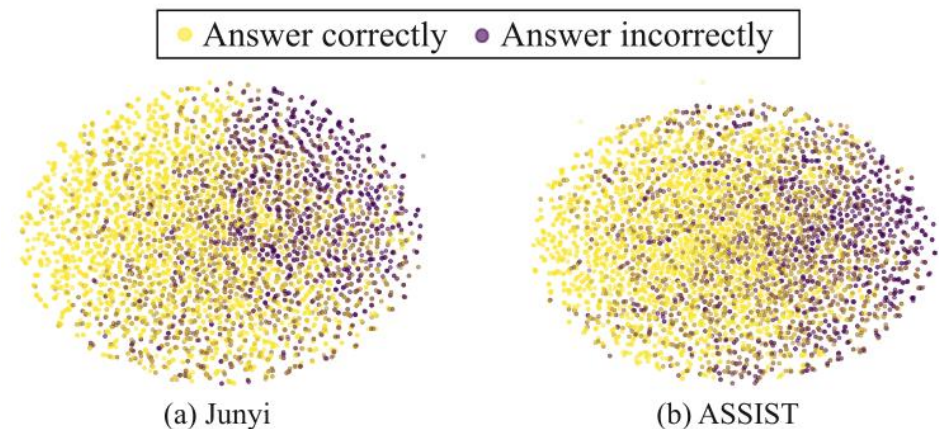
- 每个学生-练习互动关系的嵌入

$$f(r_{sq}) = f(s) - f(q)$$

- t-SNE方法将向量维度降低到二维空间
- 关系图可以捕捉到学生与练习间的交互关系

• 可解释性 (Q5)

- 同一学生有无关系图融合的诊断报告可视化
- 相似知识概念的熟练程度可以相互影响
- 沿着先决条件关系的影响在RCD中只是单向传播的



• 优势

- 对**学生-练习-概念关系**进行了全面建模，并明确地将交互和结构信息与多种关系**整合**在一起
- 具有很高的**可扩展性**，很容易改进其他认知诊断模型

• 劣势

- 概念粒度不够细致，忽略练习文本中包含的信息
- 仅根据学生在**概念上的状态**来估计学生的表现，未考虑其他影响因素（如即使一个学生没有掌握某些概念，也可能在巧合下正确地回答相关的练习）
- 未能从更深的**认知层面**对学生进行分析建模

• 未来发展

– 概念标签**细化**

- 如“数字和方程” -> “四则运算，乘法，加法”

– **动态**认知诊断

- 结合知识追踪，同时保留可解释性

– 认知过程**深度建模**

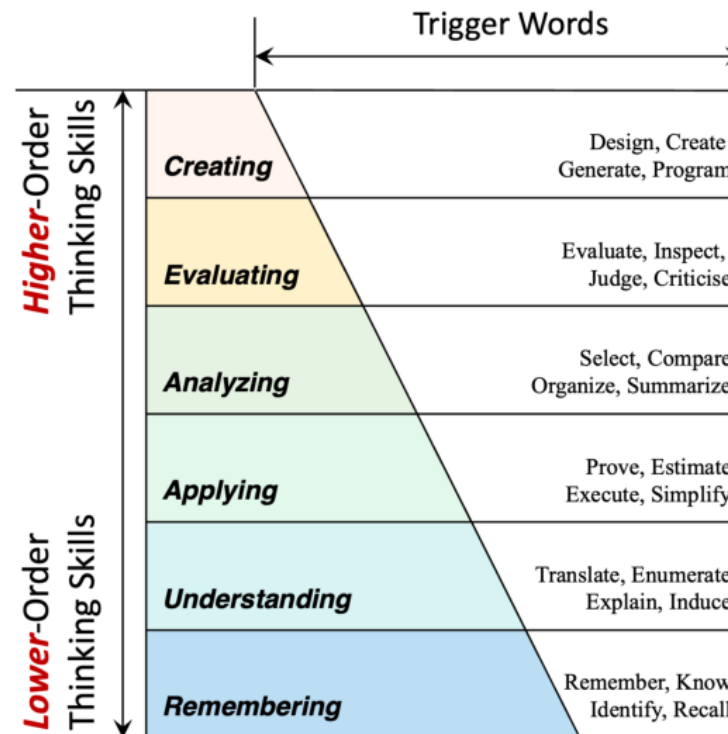
- Bloom认知理论
 - 记忆、理解、应用、分析、评价、创新

– 个性化精准认知干预/重塑

- 学习路径生成
- 学习资源/同伴推荐

E1: $2 \times 10 + 3 = \underline{\quad}$.

E2: Alice's speed is 1 m/s. John is twice as fast as Alice. John starts at 3m from the starting point of a straight runway and move forward. How far is John from the starting point after 10s?



- **Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. Neural cognitive diagnosis for intelligent education systems[C]. In Proceedings of the AAAI conference on artificial intelligence, 2020, 34: 6153–6161.**
- **Gao, W., Liu, Q., Huang, Z., Yin, Y., Bi, H., Wang, M., Ma, J., Wang, S., & Su, Y. RCD: Relation map driven cognitive diagnosis for intelligent education systems[C]. In Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval . 2021, 20: 501–510.**
- **Yu, Jifan et al. MoocRadar: A Fine-grained and Multi-aspect Knowledge Repository for Improving Cognitive Student Modeling in MOOCs. Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023, 11: 2924–2934.**

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！



• 优化Q矩阵

- 遵循贝叶斯理论, 假设 $\tilde{Q} \sim (0, \sigma)$
- 与原Q矩阵结合, 定义偏序关系

Q矩阵标注 \geq 预测top-k \geq 其他=0

$$a >_i^+ b, \text{ if } Q_{ia} = 1 \text{ and } Q_{ib} = 0 \text{ and } b \in V_i^k$$

- 定义偏序关系集

$$D_V = \{(i, a, b) | a >_i^+ b, i = 1, 2, \dots, M\}$$

- 对数后验分布

$$\ln p(\tilde{Q} | D_V) = \ln \prod_{(i,a,b) \in D_V} p(a >_i^+ b | \tilde{Q}_i) p(\tilde{Q}_i) = \sum_{i=1}^M \sum_{a=1}^K \sum_{b=1}^K I(a >_i^+ b) \ln \frac{1}{1 + e^{-\lambda(\tilde{Q}_{ia} - \tilde{Q}_{ib})}} + C - \sum_{i=1}^M \sum_{j=1}^K \frac{\tilde{Q}_{ij}^2}{2\sigma^2}$$

- 损失函数

$$\text{loss} = -\ln p(\tilde{Q} | D_V) + \text{loss}_{CDM}$$

- 认知诊断与知识追踪对比

比较维度	认知诊断 (CD)	知识追踪 (KT)
领域出发点	心理测量	人机交互
主要目的	评估学生 当前 在 知识点 层级上的能力水平	动态 预测学生中在知识点/问题上的回答正确率
特点	静态评估 ，结果精细化、多样化	动态预测 ，结果时序化
输入	学习记录、Q矩阵	学习记录、题目-知识点映射关系
输出结果	学生知识点掌握程度 预测答题情况	学生知识点掌握程度 预测答题情况 (波动性)
模型参数	可解释性	黑盒参数
应用场景	医疗诊断、 自适应测试 、个性化推荐	学习成绩预测、个性化精准推荐