

# 智慧教育中 知识追踪技术与前瞻

于戈

yuge@cse.neu.edu.cn

东北大学



# 提纲

2

- 概述
- 相关理论与模型
- 主要技术
- 未来发展方向
- 结束语



# 概述

3

- 知识追踪 (Knowledge Tracing, KT)
  - 1972年由美国心理学家Atkinson提出
  - 模拟学习者对知识的掌握情况的模型
- 定义
  - 给定一个学习者在特定学习任务上的历史交互活动序列  $X_t$ ，预测其在下一个交互活动中的学习表现与知识掌握情况
  - $X_t = \{x_1, x_2, \dots, x_t\}$ ，对于问答交互活动，学习表现  $x_{t+1} = (q_{t+1}, a_{t+1})$ ，在时间  $t + 1$ ，学习者回答了问题  $q_{t+1}$ ，得分为  $a_{t+1}$ ；
  - $K_{t+1} = (k_1, k_2, \dots, k_K)$  表示学习者在时间  $t + 1$  知识概念掌握情况， $k_i$  为学习者在时间  $t + 1$  对第  $i$  个知识点的掌握程度
  - 预测学习者在下一个时刻正确回答问题的概率与掌握知识概念的概率

$$p(a_{t+1} = 1 | q_{t+1}, X_t) \quad p(K = (1, 0, \dots, 1) | X_t)$$

# 概述

4

## □ 应用领域

- 智能教学系统 (Intelligent Tutoring System, ITS)
- 慕课系统 (Massive Open Online Course, MOOC)
- 在线评测系统 (OnlineJudge, OJ)

## □ 应用需求

- 学习者成绩预测 (Student Performance Prediction)
- 习题推荐 (Exercise Recommendations)
- 学习路径推荐 (Learning Path Recommendation)
- 自主学习服务和个性化导学

# 相关理论与模型

5

## □ 教育心理学

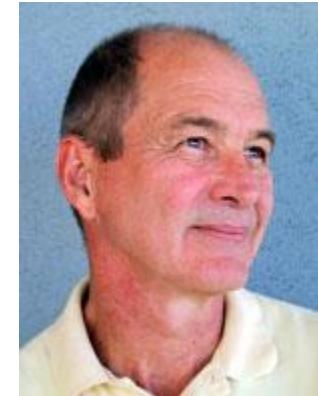
- 记忆与遗忘
- 学习风格 (learning Style)
- 项目反应理论(IRT)

## □ 机器学习

- 贝叶斯网络模型
- 因子分析模型
- 循环神经网络



艾宾浩斯  
(1850-1909)

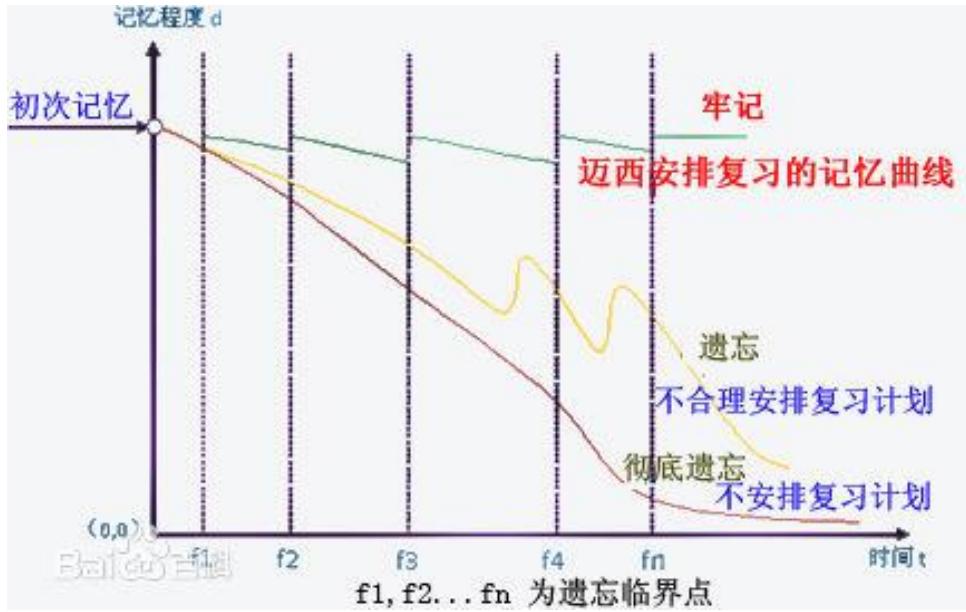


库伯  
(1939-)

# 相关理论与模型

6

## □ 教育心理学-遗忘曲线和学习风格



Ebbinghaus 遗忘曲线 (1885年)

[1] <https://baike.baidu.com/pic/遗忘曲线>

[2] <https://baike.baidu.com/item/学习圈理论>



Kolb 学习周期与学习风格 (2005年)

# 相关理论与模型

7

## □ 教育心理学-项目反应理论(IRT)

- 潜在特质模型(Latent trait model)，描述在被测试者可观察到的测试成绩与其不可观察的特质或能力之间存在的联系

## □ IRT Logistic 模型

- 单参数模型  $P(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$

- 双参数模型  $P(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$

- 三参数模型

$$P(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

$P(\theta)$ : 能力为  $\theta$  的学习者答对此题目的概率。

参数	含义
$b_i$	题目难度
$a_i$	题目区分度
$c_i$	题目猜测系数

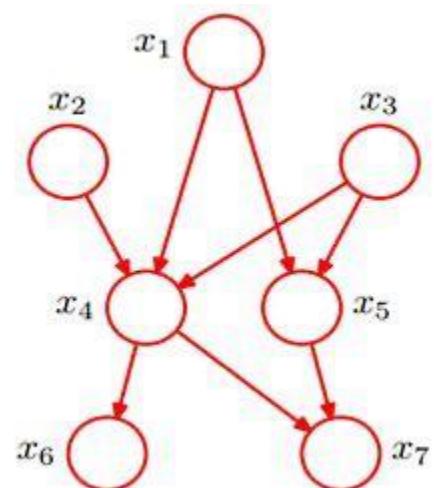
# 相关理论与模型

## □ 机器学习—贝叶斯网络(Bayesian network)

- 1988年Judea Pearl提出
- 模拟人类推理过程中因果关系的不确定性处理模型
- 贝叶斯网络=有向无环图(DAG)
  - 节点表示随机变量 $\{X_1, X_2, X_3, \dots\}$ .随机变量是可观察到的变量、隐变量、未知参数等
  - 有向边代表节点间的因果关系(由父结点指向其子结点), 用**条件概率**表达关系强度
- 基于这样的关系图获得一个概率分布



珀尔(1936-)  
2011图灵奖获得者

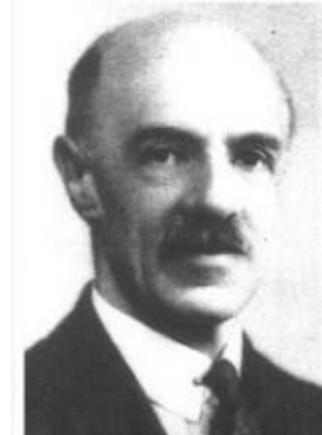


贝叶斯网络概率图

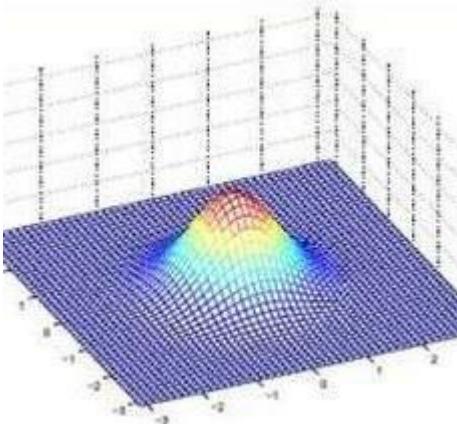
# 相关理论与模型

9

- 机器学习—因子分析(FA)
  - 从变量群中提取共性因子的统计技术
  - 提出人：C.E.斯皮尔曼
  - 现象：各科成绩之间存在一定的相关性
  - 推论：存在某些潜在的共性因子影响成绩



斯皮尔曼  
(1863-1945)



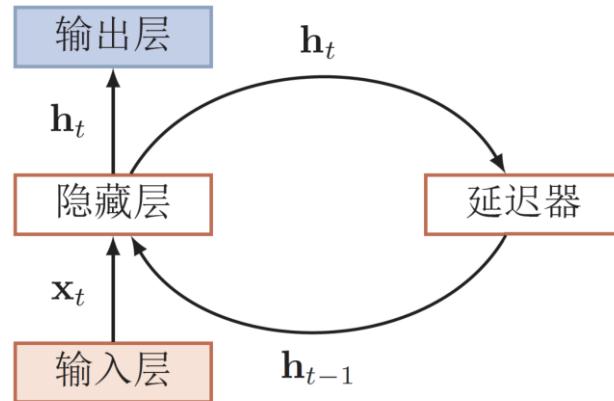
重心法、影像分析法，最大似然解、最小平方法、  
阿尔发抽因法、项目反映理论等

# 相关理论与模型

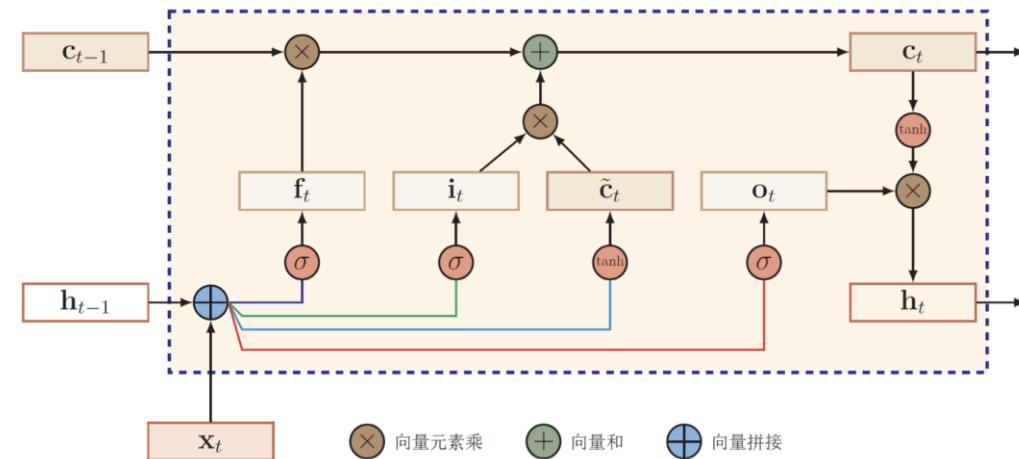
10

## □ 机器学习—循环神经网络

- 以序列数据为输入，在序列的演进方向进行递归且所有节点按链式连接的递归神经网络
- 发明于二十世纪80-90年代，并于二十一世纪初发展为深度学习算法之一



广义的循环神经网络



长短期记忆网络

# 主要技术

11

- 学习者的学习交互建模方法
- 静态模型---基于因子分析的知识追踪
- 序列模型---基于概率图模型的知识追踪模型
- 序列模型---基于深度学习的知识追踪

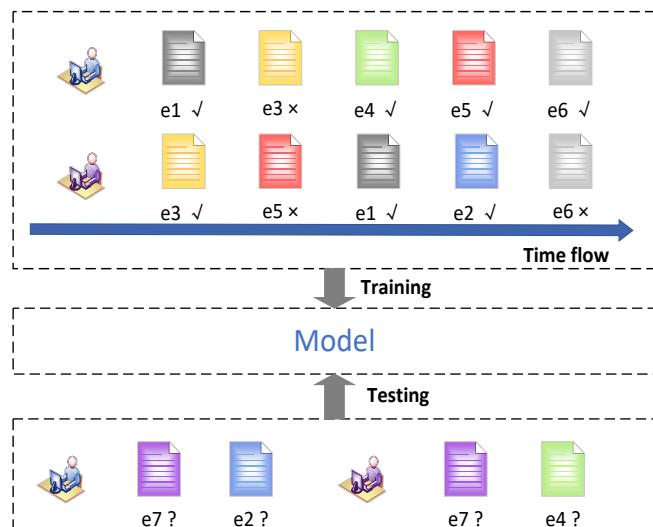
# 主要技术

12

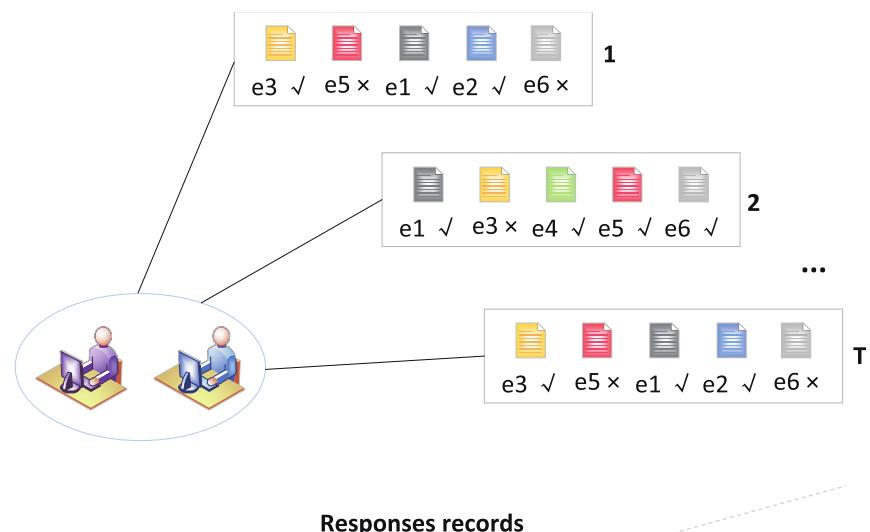
## □ 学习者学习交互的建模方法

□ 实时反馈：立刻更新模型

□ 阶段性反馈：按时间窗口更新模型



实时反馈的交互建模



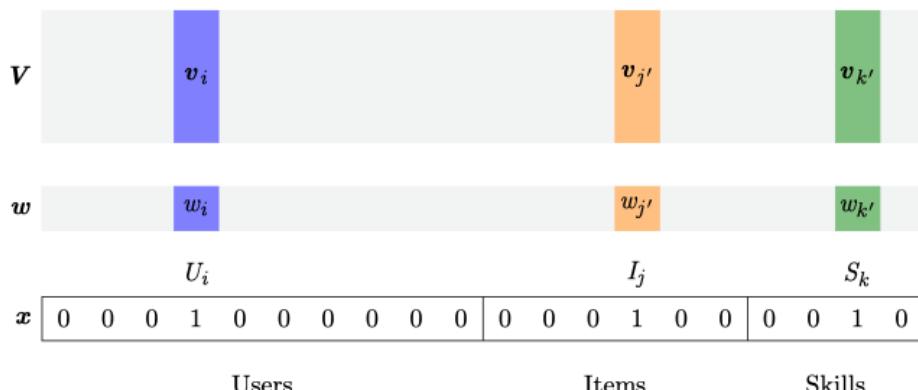
阶段性反馈的交互建模

# 主要技术

13

## □ 基于因子分析的知识追踪 (KTM)

- 将 IRT, AFM, PFA 等模型整合到一个框架下
- 挖掘了学生与题目、题目与知识概念、知识概念与知识概念之间的关系
- 将学生、题目与知识概念嵌入到相同的低维稠密空间中
- 有效的解决了稀疏数据下的特征组合问题



$$\psi(p(\mathbf{x})) = \mu + \underbrace{\sum_{k=1}^N w_k x_k}_{\text{logistic regression}} + \underbrace{\sum_{1 \leq k < l \leq N} x_k x_l \langle \mathbf{v}_k, \mathbf{v}_l \rangle}_{\text{pairwise interactions}}$$

其中  $\Psi$  为 logit 函数；  $S_k$  为题目  $I_j$  包含的知识概念；  $p(x)$  为学生  $U_i$  答对题目  $I_j$  的概率。

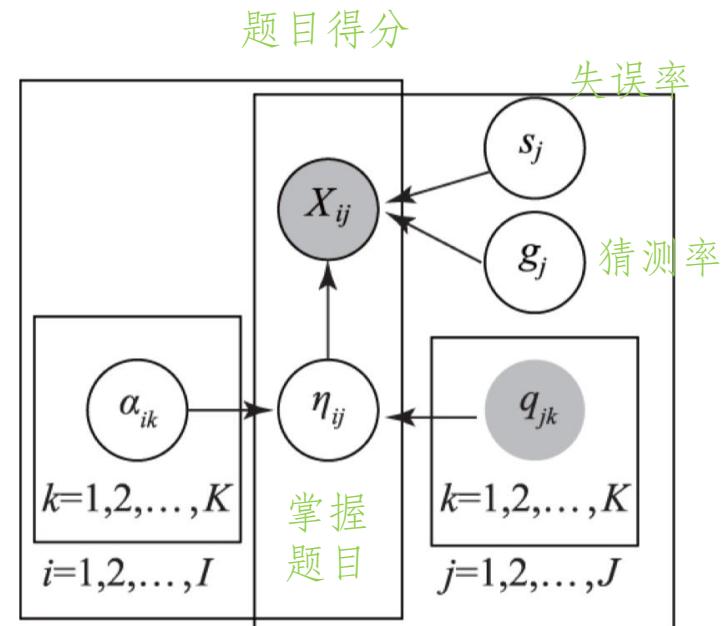
# 主要技术

14

## □ 基于概率图模型的知识追踪（DINA）

### □ DINA 模型 (Deterministic Inputs, Noisy “And” gate model)

符号	描述
$i$	学生
$j$	题目
$k$	知识点
$X$	得分矩阵
$X_{ij}$	学生 $i$ 是否答对题目 $j$
$\alpha$	潜在能力矩阵
$\alpha_{ij}$	学生 $i$ 是否掌握知识点 $j$
$\eta$	潜在作答矩阵
$\eta_{ij}$	学生 $i$ 是否掌握题目 $j$
$Q$	题目知识点关联矩阵
$q_{jk}$	题目 $j$ 对知识点 $k$ 考察情况
$s_j$	题目 $j$ 的失误率
$g_j$	题目 $j$ 的猜测率



掌握知识

题目-知识点关联

# 主要技术

15

## □ 基于概率图模型的知识追踪 (DINA)

学生*i*掌握题目*j*的概率 ( $\eta_{ij}$ 、 $\alpha_{ik}$  和  $q_{ik}$  为随机变量)

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{ik}}$$

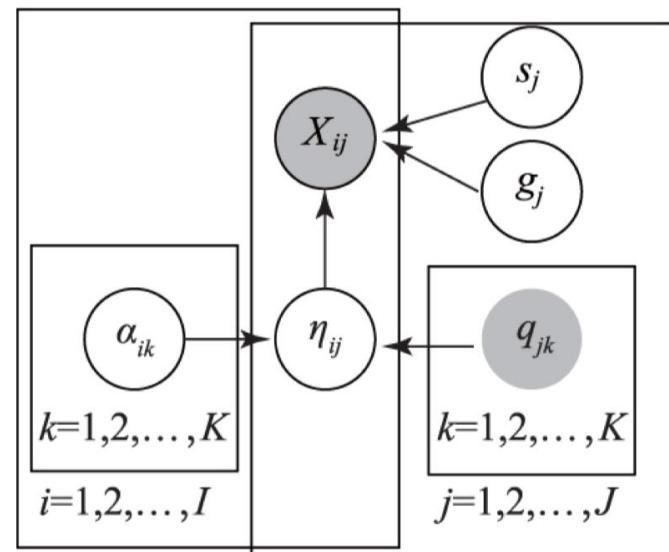
学生*i*掌握所有知识概念的概率

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

似然函数:

$$\begin{aligned} L(X) &= \prod_{i=1}^I L(x_i) \\ &= \prod_{i=1}^I \sum_{\alpha_i} P(X_i | \alpha_i) P(\alpha_i) \end{aligned}$$

其中  $\sum_{\alpha_i}$  表示变量  $\alpha_i$  的所有取值，共有  $2^k$  种情况

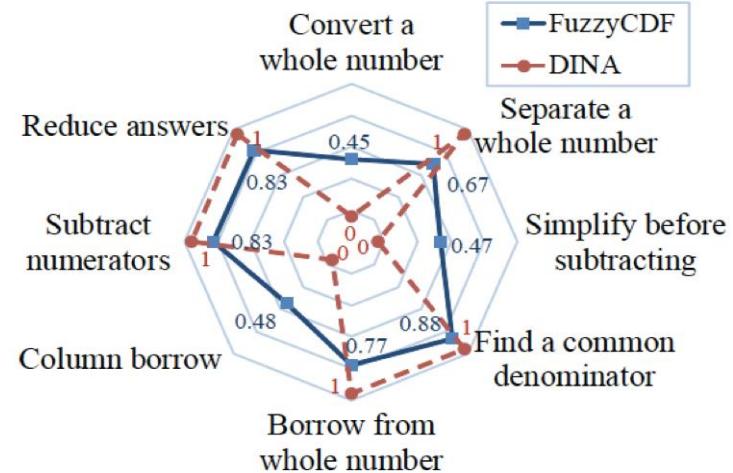
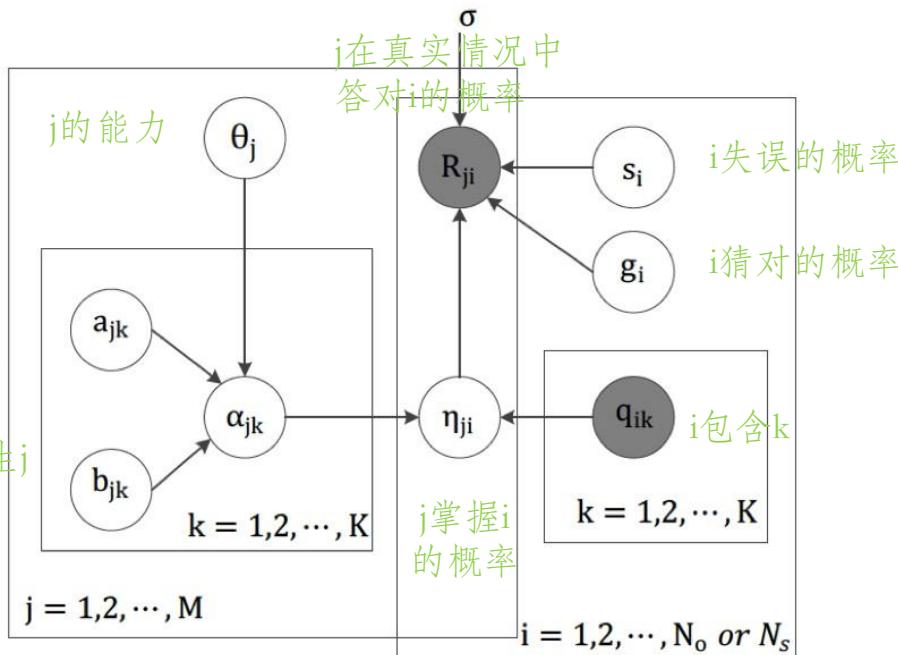


# 主要技术

16

- 基于概率图模型的知识追踪 ( FuzzyCDF )
  - 利用模糊化理论连续化学生知识点熟练度
  - 考虑主观题与客观题的区别
  - 使用蒙特卡洛马尔可夫链(MCMC)采样算法进行参数估计

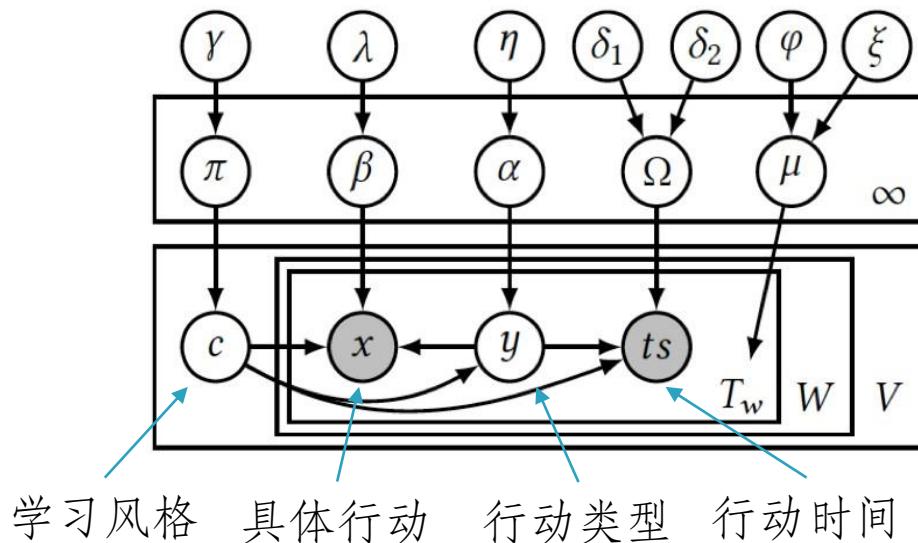
用户j  
知识点k  
题目i



# 主要技术

17

- 基于概率图模型的知识追踪 (\*)
  - 将学习风格因素融入到了知识追踪中
  - 探究了学习风格、学习类型、学习活动以及学习时间直接的关系

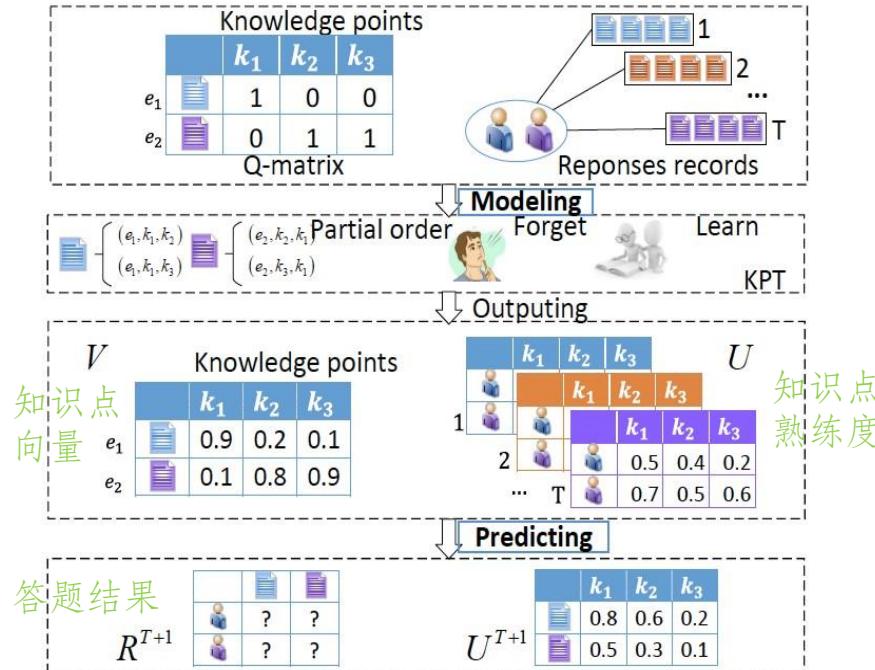
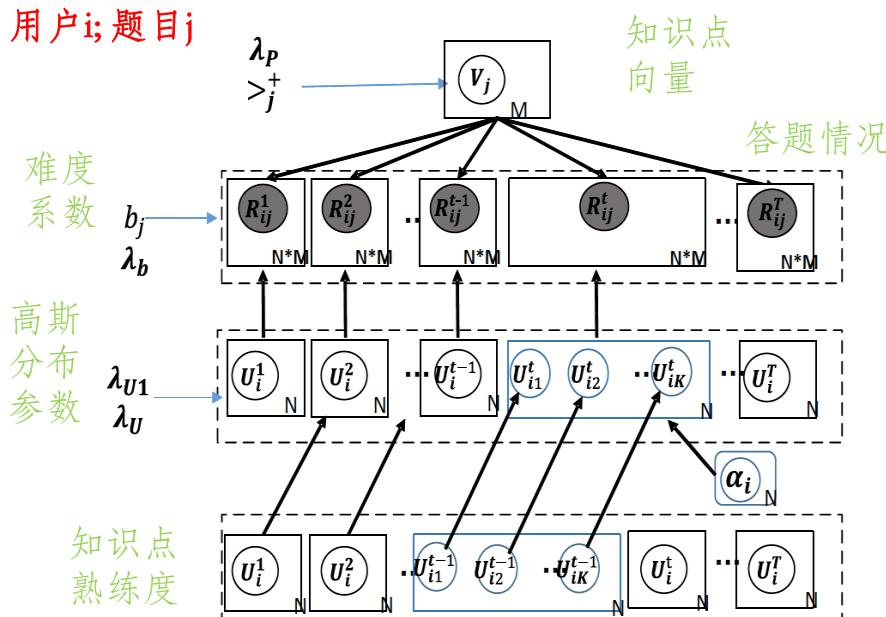


# 主要技术

18

## □ 基于概率图模型的知识追踪 (KPT)

- 利用专家标注的Q-matrix，给予模型可解释性
- 利用学习曲线与遗忘曲线作为先验，追踪学习者知识点熟练度。采用阶段性学习交互模型。



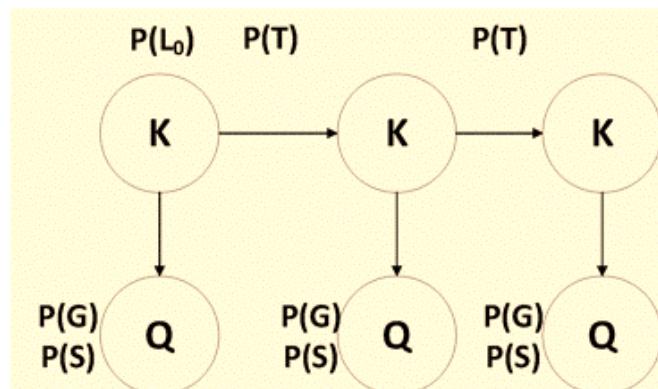
# 主要技术

19

- 基于概率图模型的知识追踪（BKT）
  - 贝叶斯知识追踪模型(BKT)
  - 采用实时反馈的学习交互模型

模型参数	模型参数对应详细解释
初始参数 $P(L_0)$	学习者尚未开始学习是知道特定知识的初始概率
学习参数 $P(T)$	经过一段时间学习之后，学习者对于知识点从不会到会的转换概率
猜测参数 $P(G)$	学习者在不知道知识点的情况下却回答正确的猜测概率
失误参数 $P(S)$	学习者在知道知识点的情况下却回答错误的失误概率
知识节点 $K$	1表示掌握，0表示未掌握
表现节点 $Q$	1表示通过，0表示未通过

知识能力



做题能力

# 主要技术

20

## □ 基于概率图模型的知识追踪 ( BKT )

状态概率矩阵		状态转移矩阵			观测矩阵		
做题前知识状态	概率	前一次知识状态	当前知识状态	概率	当前知识状态	当前表现状态	概率
已掌握知识点	$P(L)$	已掌握知识点	未掌握知识点	0	未掌握知识点	回答正确	$P(G)$
			已掌握知识点	1		回答错误	$1-P(G)$
未掌握知识点	$1-P(L)$	未掌握知识点	未掌握知识点	$1-P(T)$	已掌握知识点	回答正确	$1-P(S)$
			已掌握知识点	$P(T)$		回答错误	$P(S)$

$$p(L_1)_u^k = p(L_0)^k \quad (1)$$

$$p(L_{t+1}|Q=correct)_u^k = \frac{p(L_t)_u^k \cdot (1 - p(S))^k}{p(L_t)_u^k \cdot (1 - p(S))^k + (1 - p(L_t)_u^k) \cdot p(G)^k} \quad (2)$$

$$p(L_{t+1}|Q=wrong)_u^k = \frac{p(L_t)_u^k \cdot p(S)^k}{p(L_t)_u^k \cdot p(S)^k + (1 - p(L_t)_u^k) \cdot (1 - p(G))^k} \quad (3)$$

$$p(L_{t+1})_u^k = p(L_{t+1}|Q)_u^k + (1 - p(L_{t+1}|Q)_u^k) \cdot p(T)^k \quad (4)$$

$$p(Q_{t+1})_u^k = p(L_t)_u^k \cdot (1 - p(S))^k + (1 - p(L_t)_u^k) \cdot p(G)^k \quad (5)$$

没有考虑学习者对知识点的遗忘问题

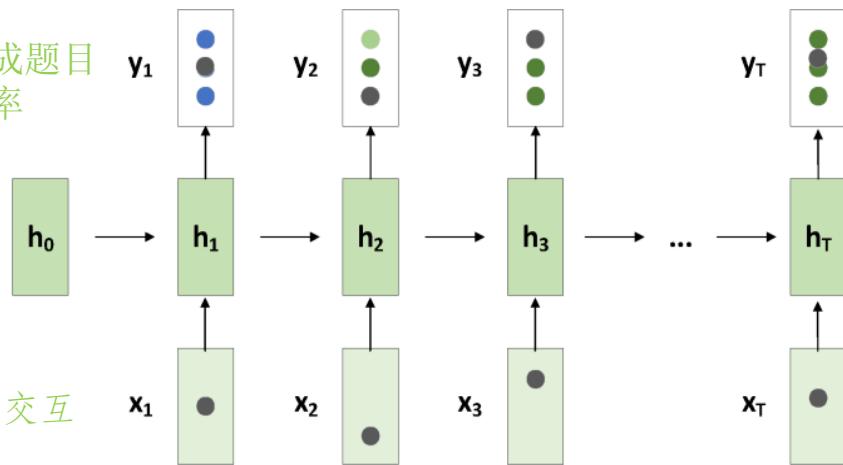
# 主要技术

21

## □ 基于深度学习的知识追踪 (DKT)

- DKT首次将RNN和LSTM应用于知识追踪
- 追踪学生知识点熟练度随时间的动态变化
- 采用实时反馈型学习交互

学生完成题目  
的正确率



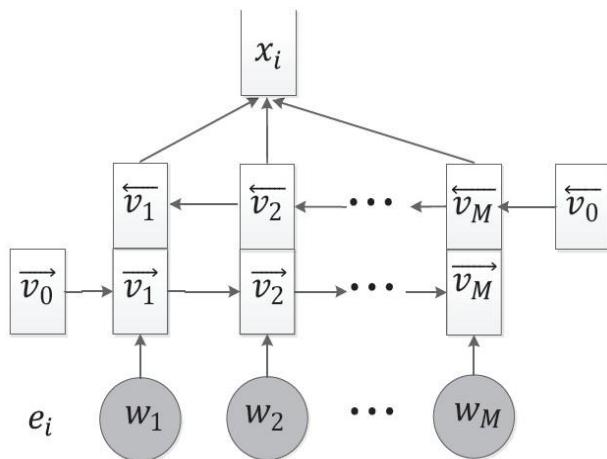
学习交互

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ C_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * C_t \\ h_t &= o_t * \tanh(C_t) \end{aligned}$$

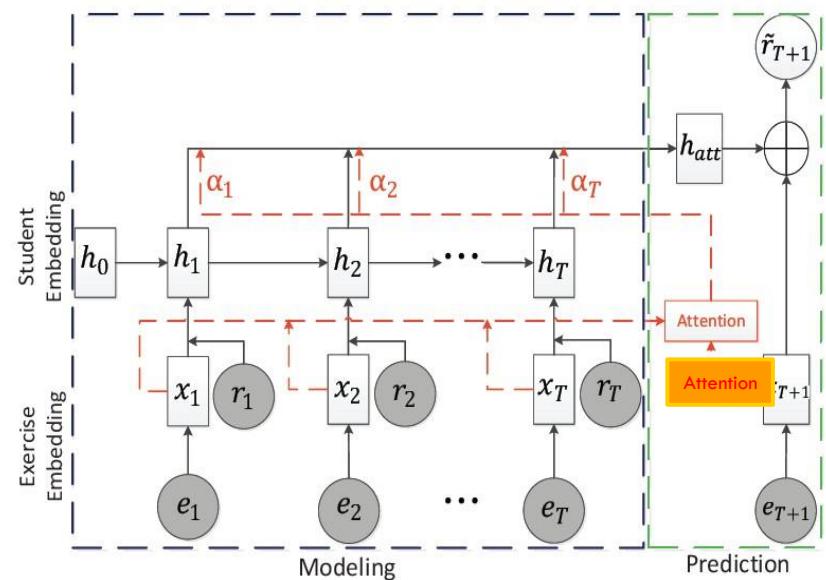
# 主要技术

22

- 基于深度学习的知识追踪 (EERNN)
  - 利用LSTM 追踪学习者的学习序列数据
  - 利用题目文本数据解决冷启动问题



练习题 $e_i$ 的文本嵌入

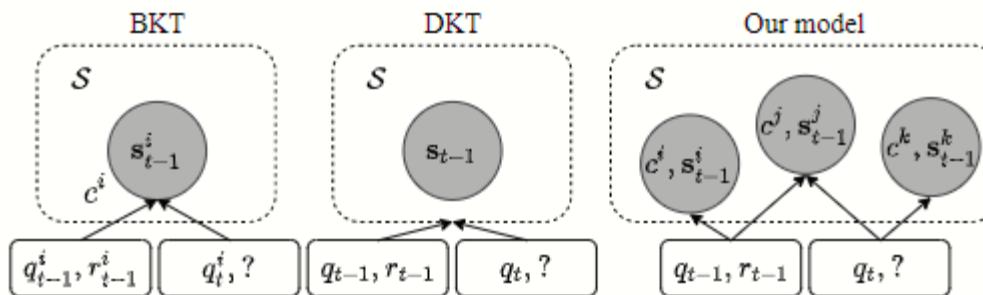


带有注意力机制的EERNN网络

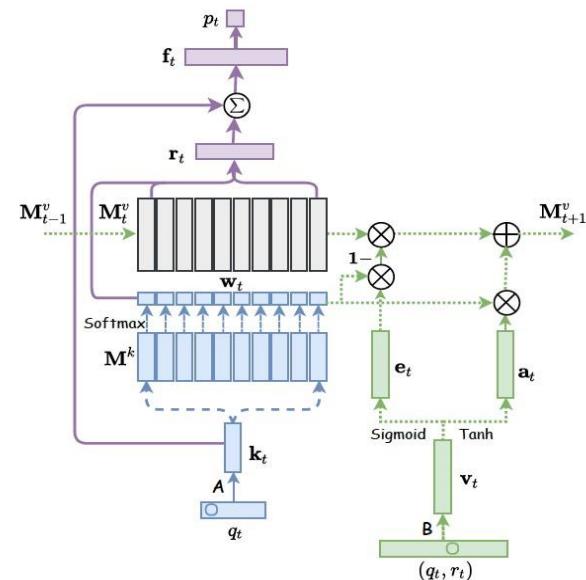
# 主要技术

23

- 基于深度学习的知识追踪 (DKVMN)
  - 引入存储模块保存知识并更新对应的知识状态
  - 自动发现通常由人类专家标注的概念
  - 描述学生变化的知识状态



动态键值存储网络与BKT、DKT的区别

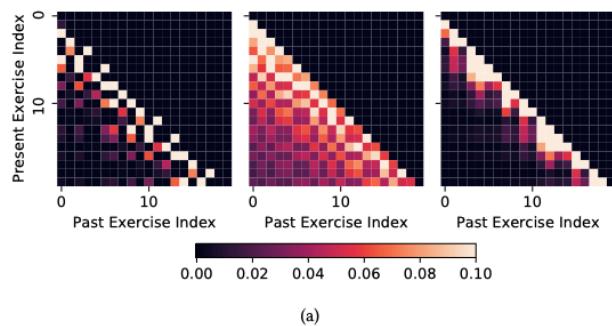


动态键值存储网络

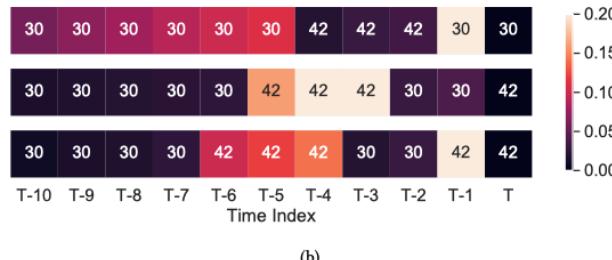
# 主要技术

24

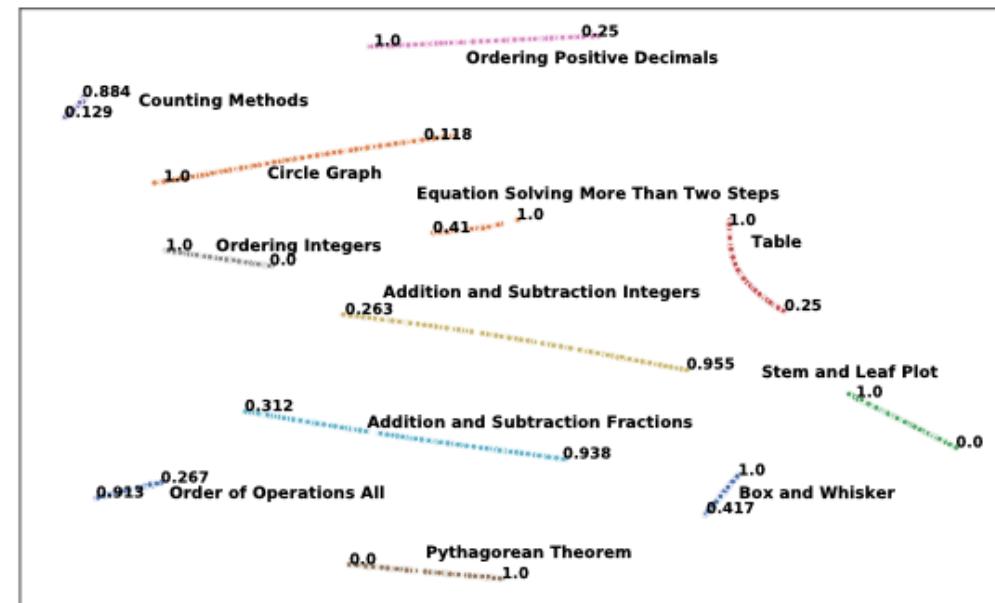
- 基于深度学习的知识追踪 (AKT)
  - 基于 Rasch 模型的题目与知识概念嵌入
  - 基于上下文的题目与知识概念编码
  - 利用单调注意力机制建模学生学习过程中的遗忘



(a)



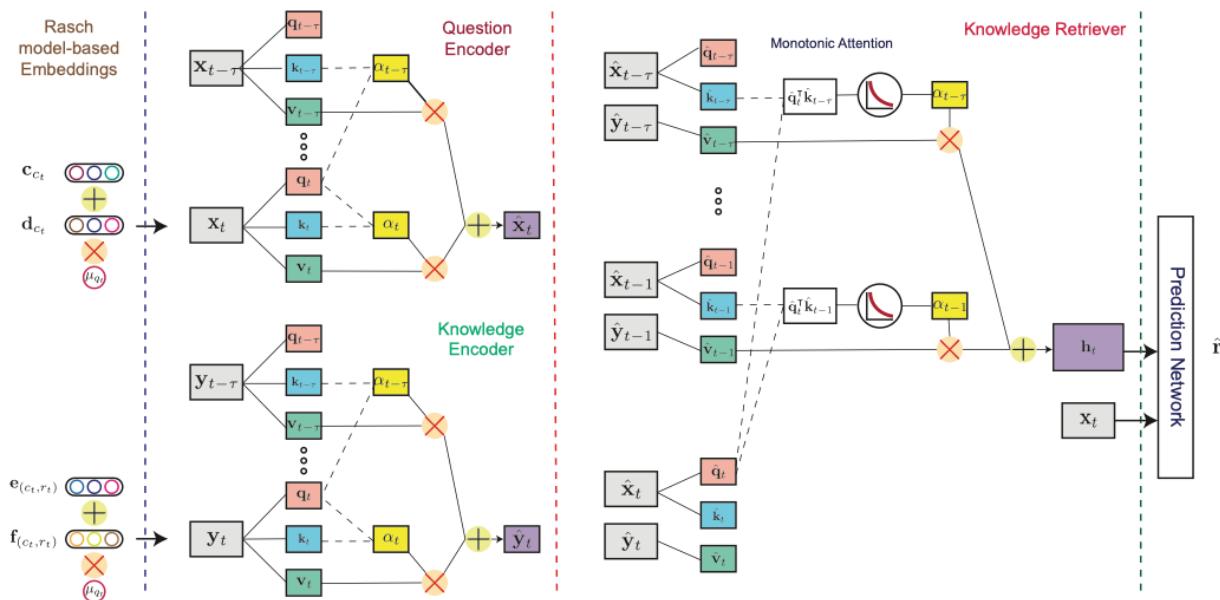
(b)



# 主要技术

25

- 基于深度学习的知识追踪 (AKT)
  - 基于 Rasch 模型的题目与知识概念嵌入
  - 基于上下文的题目与知识概念编码
  - 利用单调注意力机制建模学生的记忆与遗忘



# 方法对比

26

## □ 现有知识追踪方法对比

- 仅考虑学习的序列数据，缺乏对于其他领域数据的利用
- 仅对单一因素的挖掘，缺乏一个通用的模型

Model	数据源					预测		解释性
	Q-matrix	多知识点	重复练习	文本	时间	多知识点	学生反应	
IRT	x	x	x	x	x	x	✓	✓
DINA	✓	✓	x	x	x	✓	✓	✓
KTM	✓	✓	✓	x	x	x	✓	x
BKT	✓	x	✓	x	✓	✓	✓	✓
FuzzyCDF	✓	✓	x	x	x	✓	✓	✓
KPT	✓	✓	x	x	✓	✓	✓	✓
DKT	x	x	✓	x	x	x	✓	x
EERNN	x	x	✓	✓	x	x	✓	x
DKVMN	x	x	✓	✓	x	x	✓	x
AKT	x	✓	✓	x	✓	x	✓	✓

# 应用举例：OnlineJudge 知识追踪

27

## □ 东北大学在线评测系统(NEUOJ)

- 基于学习迁移理论的BKT方法
- 知识点关联度分析
- 学生知识熟练度度量
- 考察过去和今后动态变化情况



HOME PROBLEM STATUS CONTEST TRAINING DISCUSS RATING HACKME

### Contest List

Contest ID	Contest Title	Start Time	Type	Status	Ranklist
170	CSE-2019-CPP-继承&多态&模板	2019-07-10 19:30:00	Private	Running	Watch
169	CSE-2019-CPP-类	2019-07-10 19:30:00	Private	Running	Watch
168	2019.5? monthly	2019-06-09 12:00:00	Public	Ended	Watch
167	东北大学 2019 年“建龙钢铁”大学生...	2019-05-13 20:00:00	Private	Ended	Watch
166	TEST_东北大学2019年“建龙钢铁”大...	2019-05-12 01:00:00	Private	Ended	Watch
165	2019东北大学校赛重现	2019-05-04 13:00:00	Public	Ended	Watch
164	2019.4 monthly	2019-05-01 12:00:00	Public	Ended	Watch
163	2019上半年-选修课04	2019-04-27 00:00:00	Public	Ended	Watch
162	2019年“建龙钢铁”大学生程序设计竞...	2019-04-21 14:00:00	Private	Ended	Watch
160	2019上半年-选修课03	2019-04-20 19:00:00	Public	Ended	Watch
159	neu_test training	2019-04-16 16:50:00	Private	Ended	Watch
158	2019上半年-选修课02	2019-04-13 16:00:00	Public	Ended	Watch
157	校赛输入数据规范说明	2019-04-09 20:50:00	Public	Ended	Watch
156	2019上半年-选修课01 (选修课名单)	2019-03-30 15:30:00	Private	Ended	Watch
155	2019.2 monthly	2019-02-17 12:00:00	Public	Ended	Watch

# 未来发展方向

28

## □ 存在的问题

- 多模态内容方面：现有的知识追踪模型没有能力处理教学过程中视频、图像等信息。
- 认知结构方面：现有的知识追踪模型由于缺乏对于学习者认知结构的了解，无法追踪学习者与知识点之间更深层次的关系。
- 可解释性方面：现有基于深度学习的知识追踪模型注重预测学生未来表现，但是其中间结果不具有可解释性，不能够有效的展现学生对各个知识概念的掌握程度。
- 合作学习方面：现有的知识追踪模型多是对单个学习者对于知识掌握预测，缺少了小组学习、合作学习的学习者知识掌握的研究。

# 未来发展方向

29

- 深度学习
  - 使用多模态深度学习技术来有效的利用教学过程中的多模态数据
  - 基于图神经网络(GNN)的知识追踪，考虑隐性的知识图结构
  - 增强模型的可解释性，可以有效的追踪学生各个时刻对知识概念的掌握程度。
- 教育心理学理论
  - 结合教育心理学理论，设计学习者的认知结构模型
- 交互学习空间
  - 面向小组学习、合作学习/协作学习的知识追踪

# 结束语

- 知识追踪是教育科技领域（EduTech）的关键技术之一
- 对于提高教学水平、学习能力和教育质量，具有重要作用
- 目前仍处于探索阶段和初级应用，对于解决高阶知识的追踪仍有许多问题需要探索

# 结束语

请提宝贵意见

谢谢！

刘恒宇,张天成,武培文,于戈.知识追踪综述[J].华东师范大学学报(自然科学版),2019(05):1-15.