



南宁师范大学学术报告会

基于学习迁移理论的知识 结构与熟练度追踪方法

于戈 刘恒宇

东北大学计算机学院

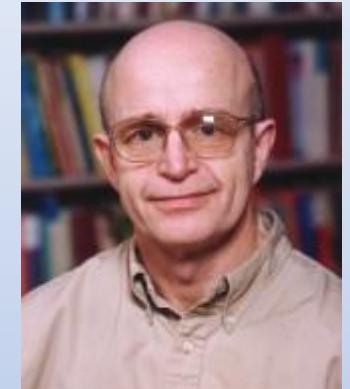
2022.11.8

提纲

1. 知识追踪原理
2. 知识追踪主要技术
3. 教育心理学中的学习迁移理论
4. 基于学习迁移的知识结构追踪
5. 建模方法与实验评价
6. 总结

知识追踪原理

- 知识追踪（Knowledge Tracing, KT）
 - 追踪学习者对知识的掌握情况。1972年由美国心理学家 Atkinson提出概念。认知心理学与计算机科学的交叉研究
 - 1995年美国CMU的Corbett博士和Anderson教授最早建立BKT 模型和算法
- 什么是知识
 - 一般认为，知识是事实、信息的描述或在教育和实践中获得的技能
- 认知心理学：
 - **陈述性知识：**描述客观事物的特点及关系，包括符号表征、概念、命题
 - **程序性知识：**也称操作性知识。关于办事的操作步骤，“做什么”和“如何做”



知识追踪原理

- 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 $x_{t+1} = (q_{t+1}, r_{t+1})$, 在时间 $t + 1$, 学习者回答了问题 q_{t+1} , 得分为 r_{t+1} ;
- $U_{t+1} = (k_1, k_2, \dots, k_K)$ 表示学习者在时间 $t + 1$ 知识掌握情况, 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的概率

$$p(r_{t+1} = 1 | q_{t+1}, X_t) \text{ 和 } p(U = (1, 0, \dots, 1) | X_t)$$

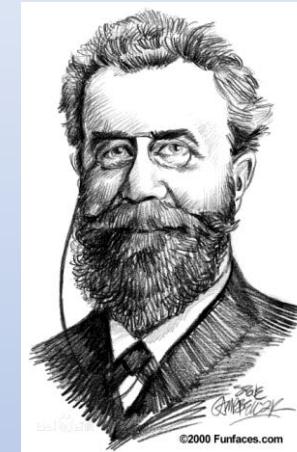
知识追踪应用

- 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 System, 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
 -

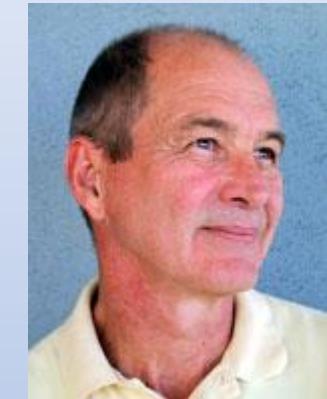


相关理论与模型

- 教育学理论
 - 记忆与遗忘、学习风格 (Learning Style)
- 教育心理学
 - 项目反应理论(IRT)、认知诊断模型
- 数据挖掘
 - 贝叶斯网络、概率矩阵分解
- 人工智能
 - 学习者模型、深度学习技术



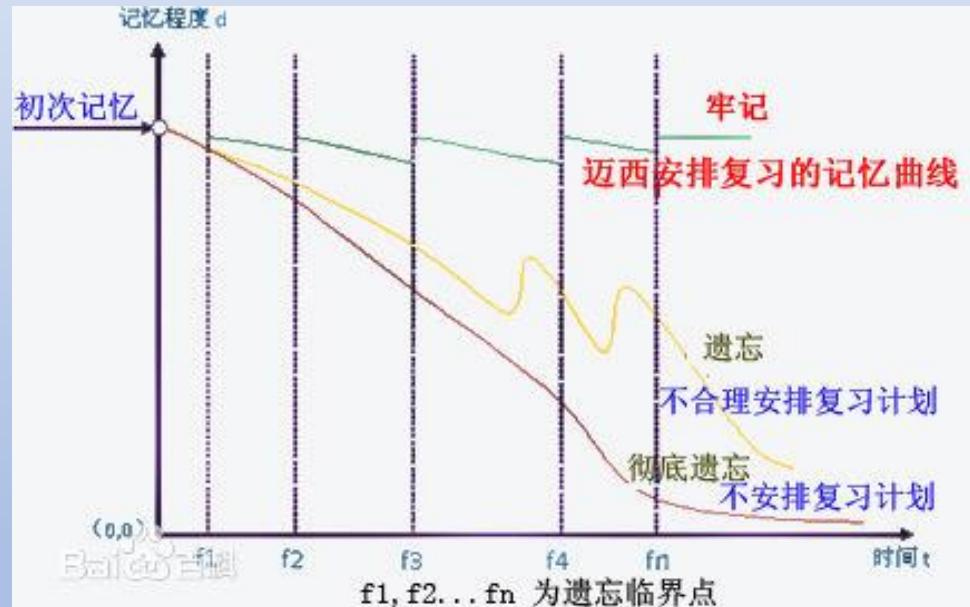
艾宾浩斯
(1850-1909)



库伯
(1939-)

相关理论与模型

- 教育学理论-记忆曲线，遗忘曲线以及学习风格



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

- 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

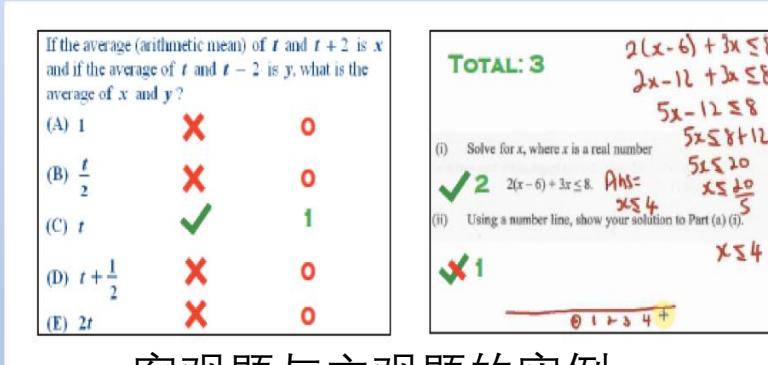
$$P(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

$P(\theta)$: 能力为 θ 的学习者答对此题目的概率。

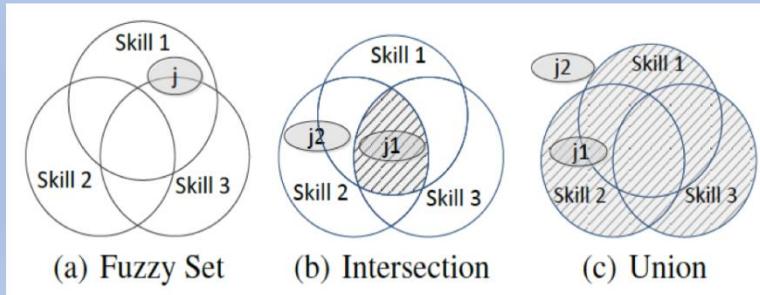
参数	含义
b_i	题目难度
a_i	题目区分度
c_i	题目猜测系数

相关理论与模型

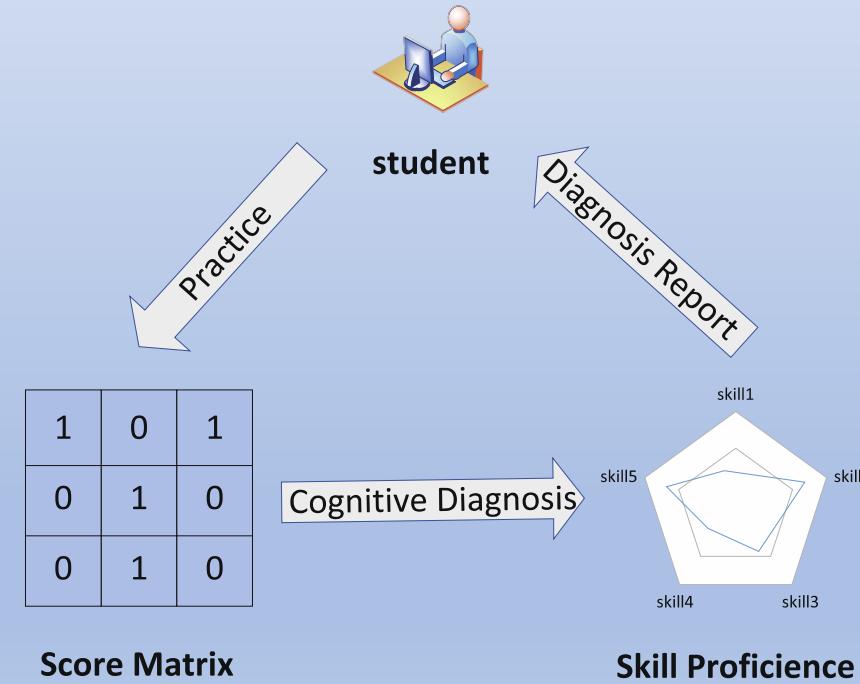
- 教育心理学--认知诊断模型 (CDM)
 - 对学习者的认知过程、加工技能或知识结构的诊断评估方法
 - 探索学习者行为、知识结构与题目之间的关系



客观题与主观题的实例



对于知识点掌握能力的不同考察方式



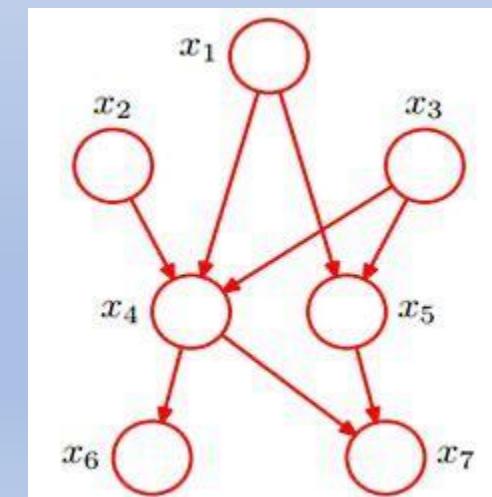
认知诊断过程

相关理论与模型

- 数据挖掘理论--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
 - 观测结点表示观测到的数据, 隐含结点表示潜在的知识, 边描述知识与数据的相互关系



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

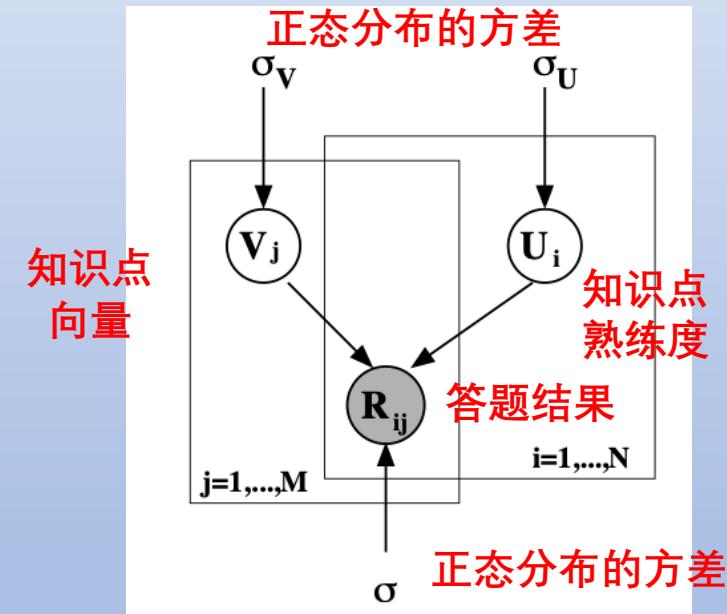
- 数据挖掘理论--概率矩阵分解(PMF)
 - 来源于个性化推荐的协同过滤算法
 - 预测出学习者成绩矩阵中的缺失值
 - 可以处理大型、稀疏且不平衡的数据集

$$p(R|U, V, \sigma^2) = \prod_{i=1}^N \prod_{j=1}^M \left[\mathcal{N}(R_{ij} | U_i^T V_j, \sigma^2) \right]^{I_{ij}}$$

$$p(U|\sigma_U^2) = \prod_{i=1}^N \mathcal{N}(U_i | 0, \sigma_U^2 \mathbf{I})$$

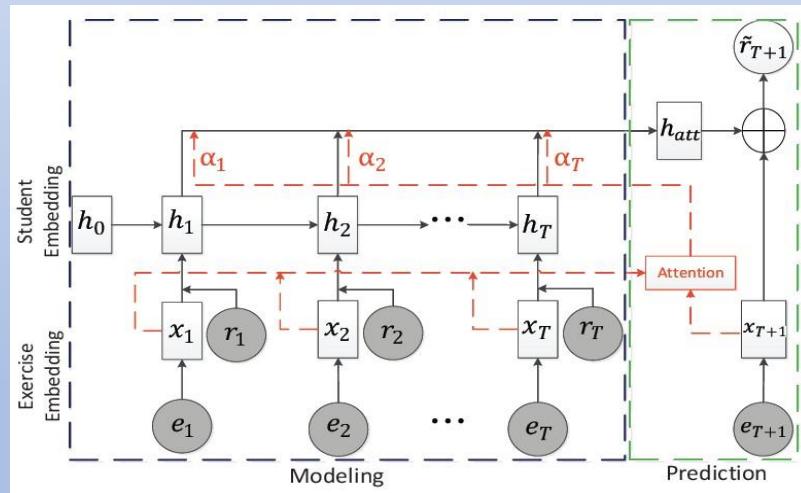
$$p(V|\sigma_V^2) = \prod_{j=1}^M \mathcal{N}(V_j | 0, \sigma_V^2 \mathbf{I})$$

$$\ln p(U, V|R, \sigma^2, \sigma_V^2, \sigma_U^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - U_i^T V_j)^2 - \frac{1}{2\sigma_U^2} \sum_{i=1}^N U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^M V_j^T V_j$$

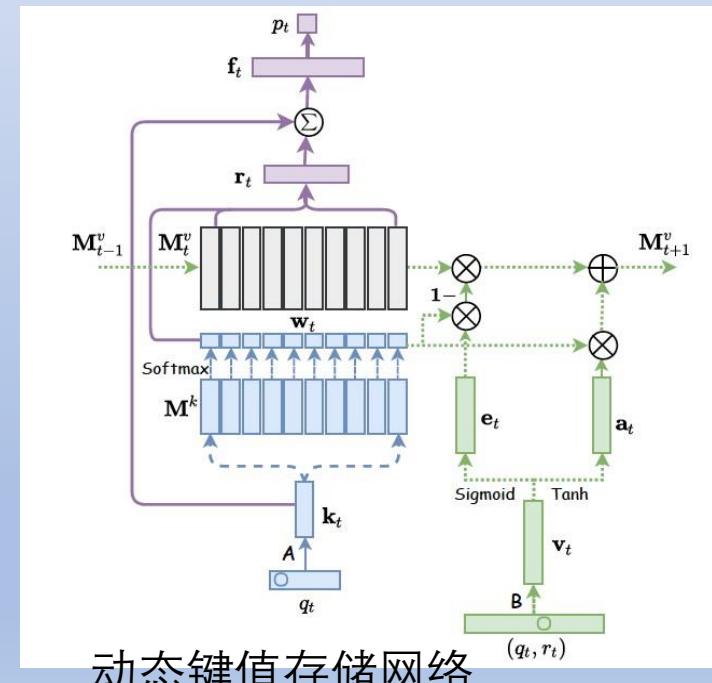


相关理论与模型

- 人工智能理论--深度学习
 - 处理学习者的学习序列数据，预测学习者的未来成绩，追踪学习者知识熟练度的动态变化
 - 使用充分的文本数据来解决冷启动问题



带有注意力机制的EERNNA网络



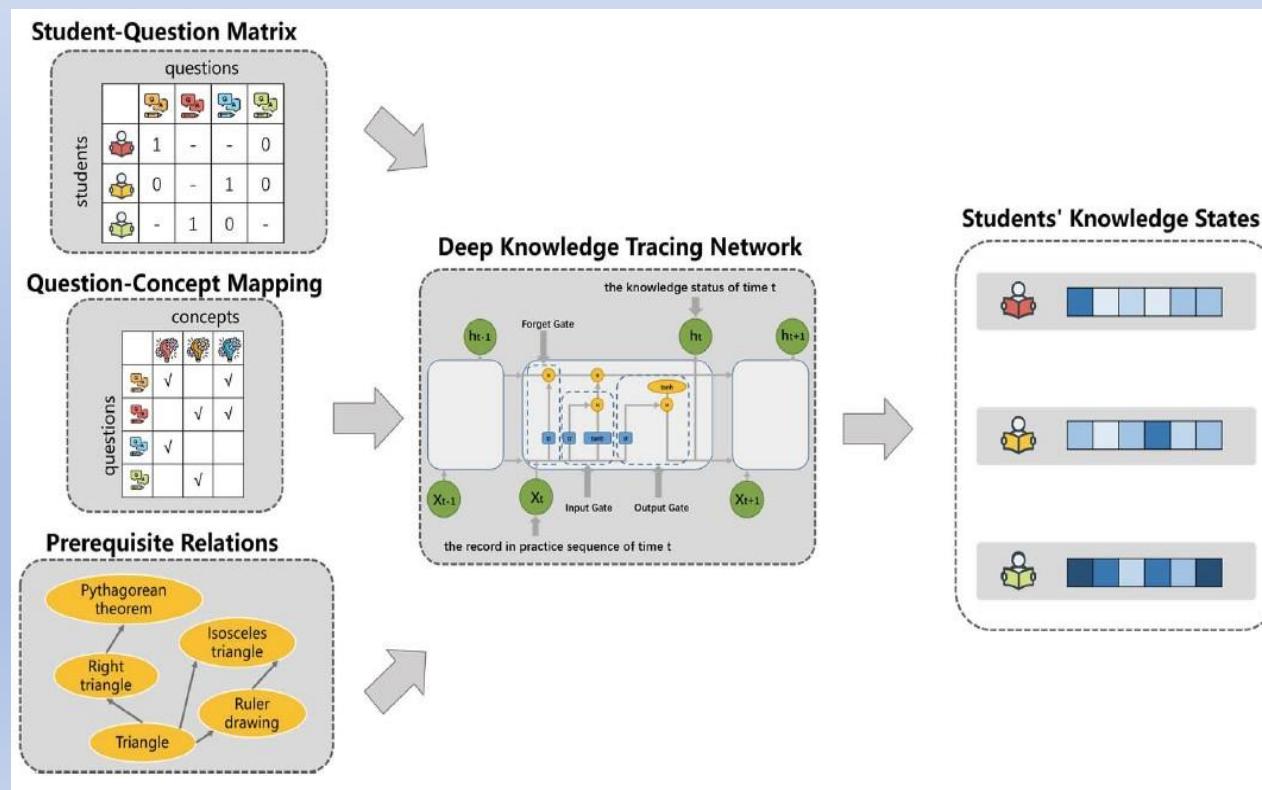
动态键值存储网络

- [1] Su Y, Liu Q, Liu Q, et al. Exercise-enhanced sequential modeling for student performance prediction[C]. Proc. 32nd AAAI Conf., 2018.
[2] Zhang J, Shi X, King I, et al. Dynamic key-value memory networks for knowledge tracing[C]. Proc. Od 26th WWW¹², 2017: 765-774.

相关理论与模型

- 人工智能理论--学习者模型

- 对学习者做题的行为进行建模，包括：学习者练习记录，题目知识概念关系和知识概念关系等
- 对学习者复杂的认知状态和知识结构进行准确描述与量化



[1] Chen P, Lu Y, Zheng V W, et al. Prerequisite-driven deep knowledge tracing[C]//2018 IEEE ICDM, 2018: 39-48.

知识追踪主要技术

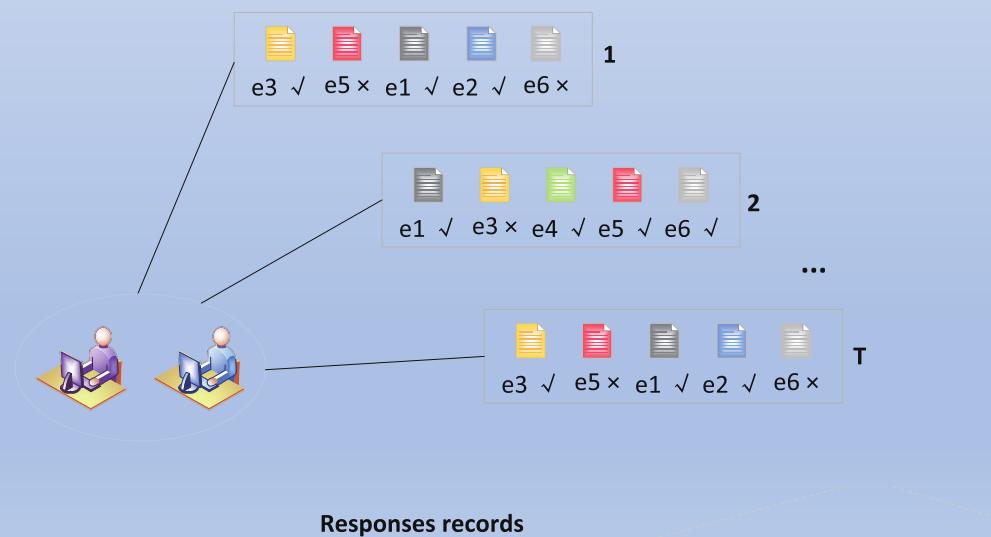
- 学习者的学习交互建模方法
- 基于概率图模型的知识追踪
- 基于矩阵分解的知识追踪
- 基于深度学习的知识追踪

知识追踪主要技术

- 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



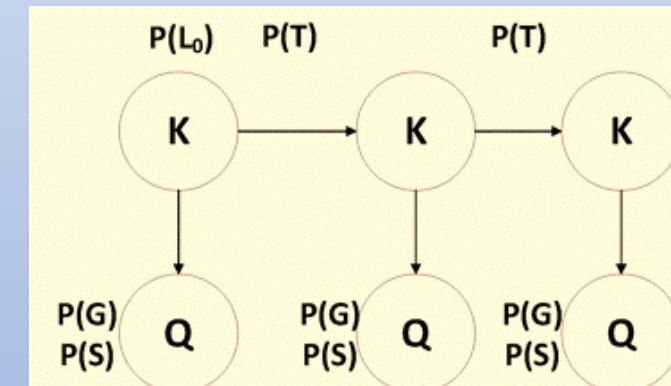
阶段性反馈的交互建模

知识追踪主要技术

- 基于概率图模型的知识追踪（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

知识追踪主要技术

- 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	$P(S)$

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

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

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

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

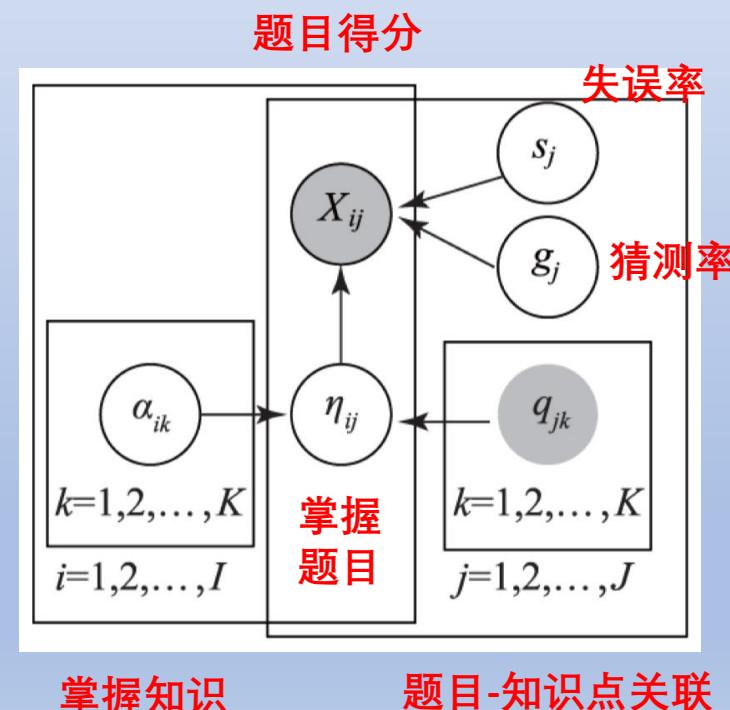
$$p(Q_{t+1})_u^k = p(L_t)_u^k \cdot (1 - p(S))^k + (1 - p(L_t)_u^k) \cdot p(G)^k \quad (5)$$

没有考虑学习者
对知识点的遗忘问题

知识追踪主要技术

- 基于概率图模型的知识诊断 (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>i</i> 是否答对题目 <i>j</i>
α	潜在能力矩阵
α_{ij}	学生 <i>i</i> 是否掌握知识点 <i>j</i>
η	潜在作答矩阵
η_{ij}	学生 <i>i</i> 是否掌握题目 <i>j</i>
Q	题目知识点关联矩阵
q_{jk}	题目 <i>j</i> 对知识点 <i>k</i> 考察情况
s_j	题目 <i>j</i> 的失误率
g_j	题目 <i>j</i> 的猜测率



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

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

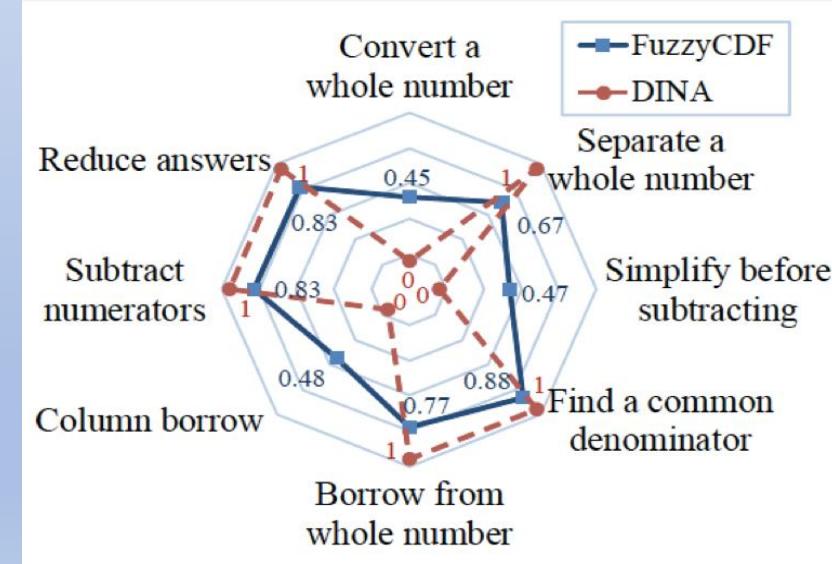
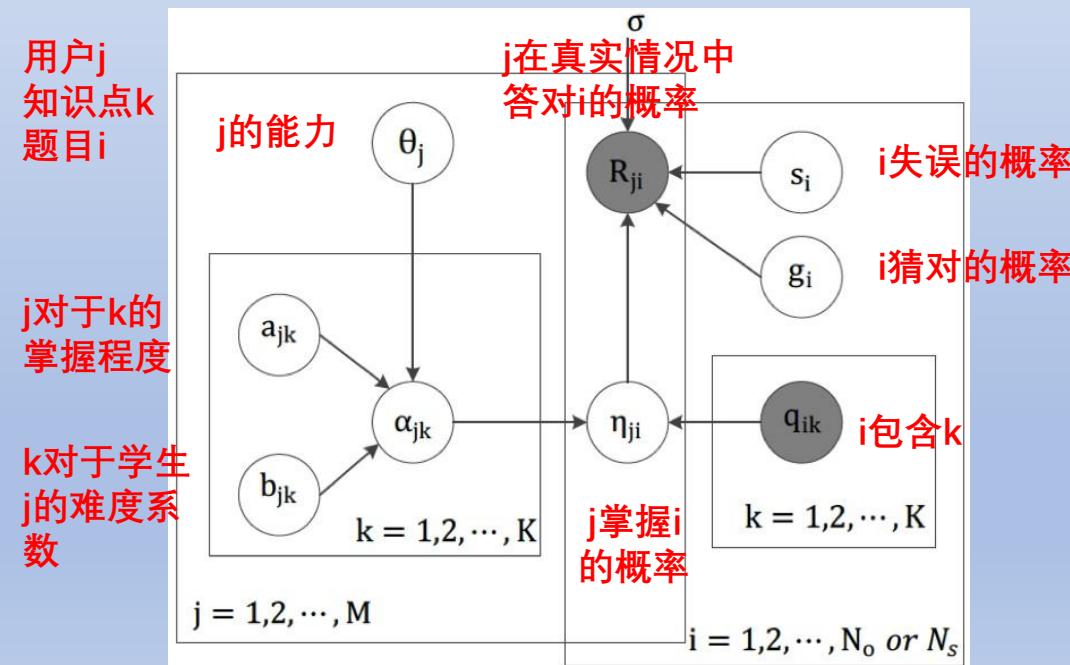
似然函数:

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

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

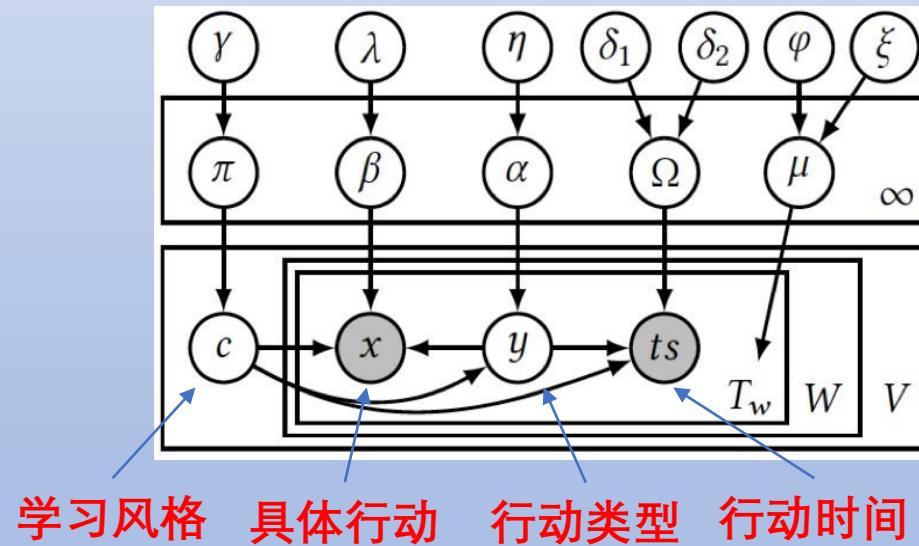
知识追踪主要技术

- 基于概率图模型的知识追踪（模糊认知诊断模型 - 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



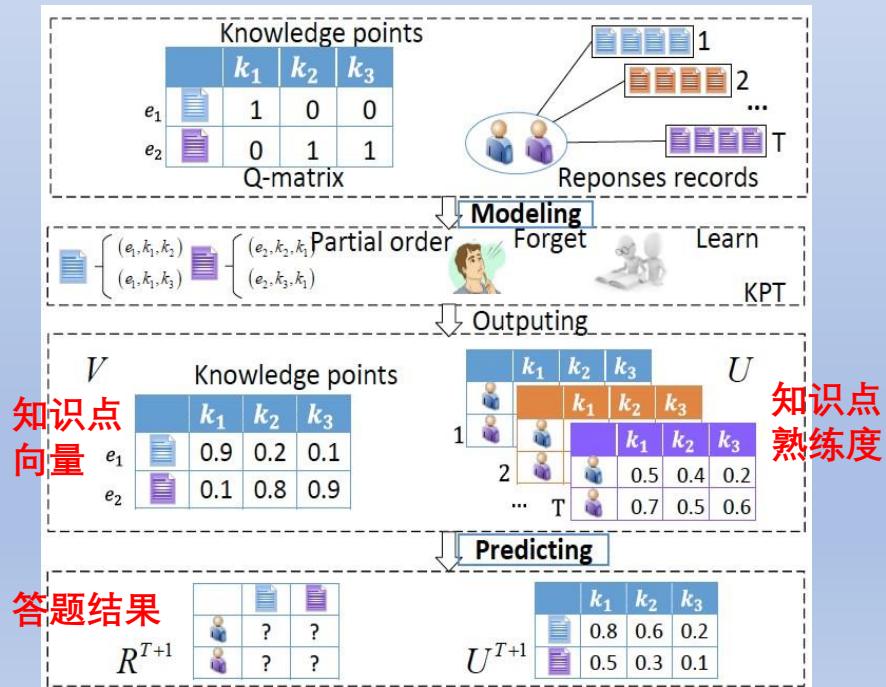
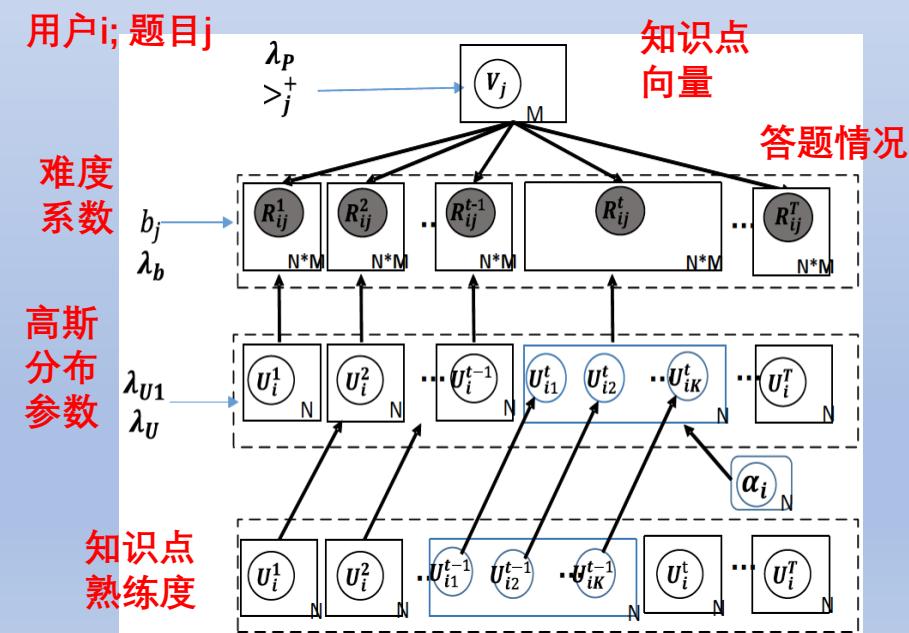
知识追踪主要技术

- 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



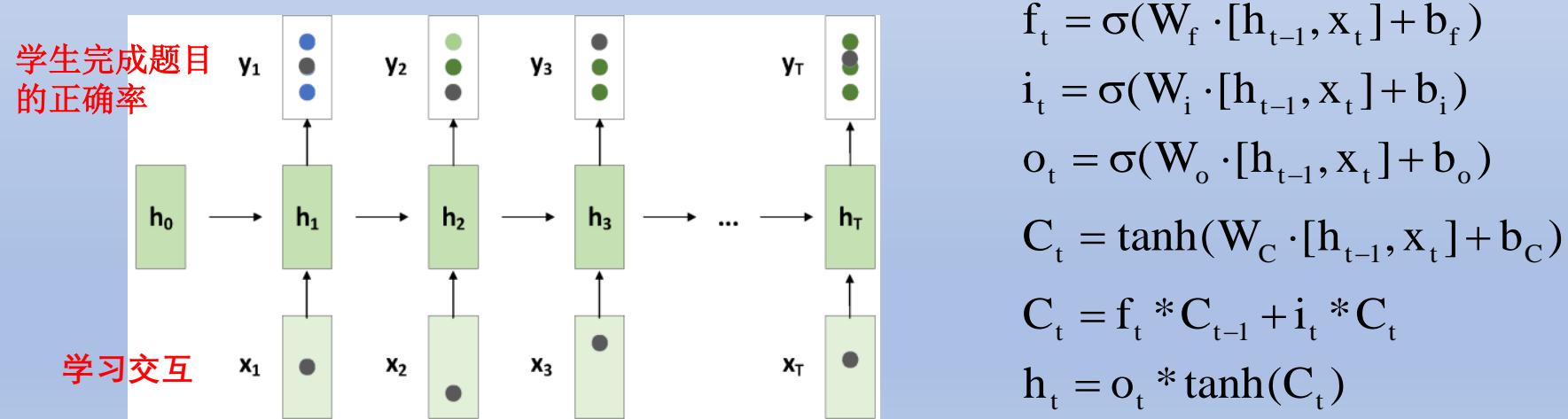
知识追踪主要技术

- 基于矩阵分解的知识追踪 (知识点熟练度追踪方法 - 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。



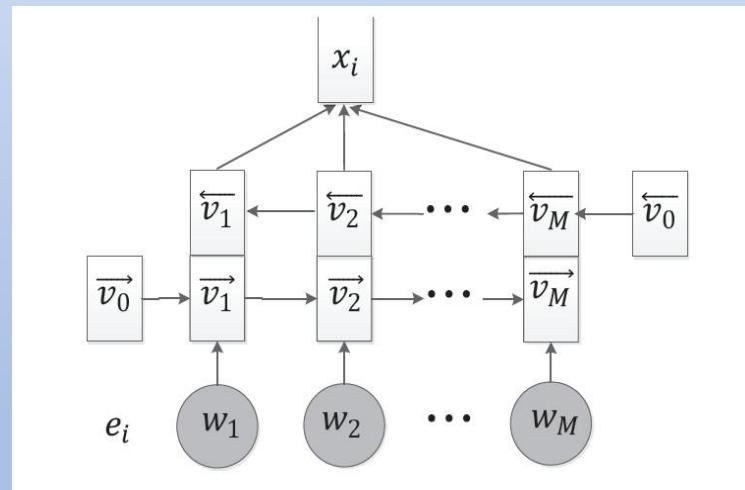
知识追踪主要技术

- 基于深度学习的知识追踪(深度知识追踪方法 - 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
 - 不足：每到题目仅仅包含一个知识点

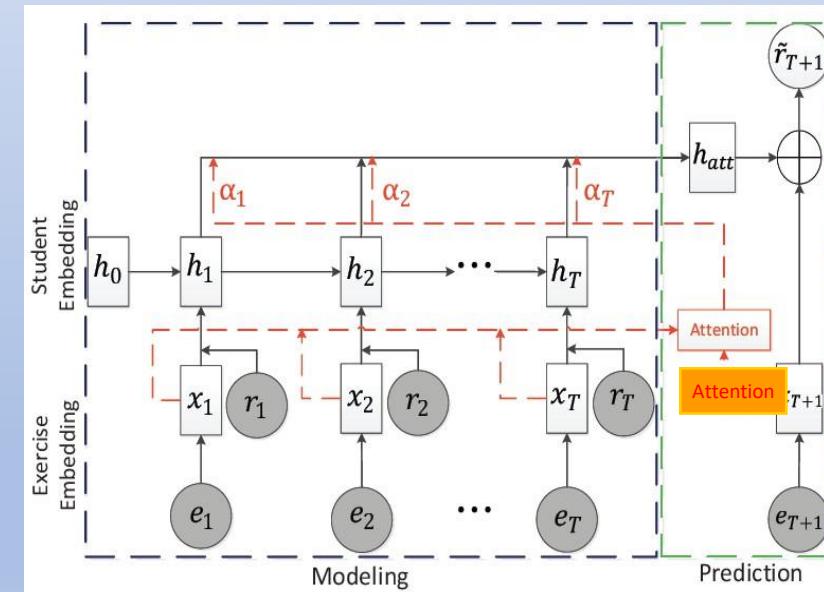


知识追踪主要技术

- 基于深度学习的知识追踪(DKT)
 - 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EERNNA网络

知识追踪面临的问题与挑战

着重追踪知识概念熟练度的动态变化

缺少追踪学习者的知识结构的动态变化

缺少学习迁移现象对于学习者学习的影响

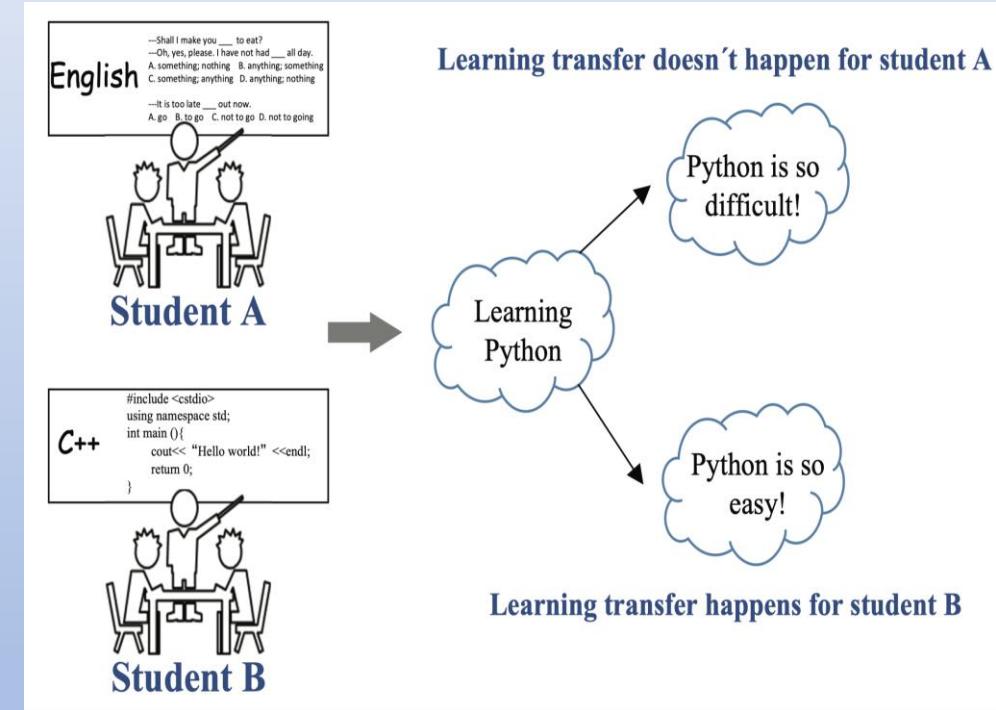
教育学中的学习迁移理论

• 概述

- 一种学习对另一种学习的影响
- "为迁移而教，为迁移而学"

• 基本分类

- **正迁移**: 一种学习对另一种学习产生的促进作用
- **负迁移**: 一种学习对另一种学习产生的干扰作用
- **顺向迁移**: 先行学习对后继学习的影响
- **逆向迁移**: 后继学习对先行学习的影响



认知结构迁移理论

- 奥苏贝尔于1963年提出
- 学习迁移主要受到一下因素影响
 - 原有知识结构的可利用性
 - 对于知识抽象原理的理解
 - 原有知识结构的稳固性
 - 知识点的掌握程度
 - 原有知识结构的可辨别性
 - 学习者对于知识点之间关系的理解



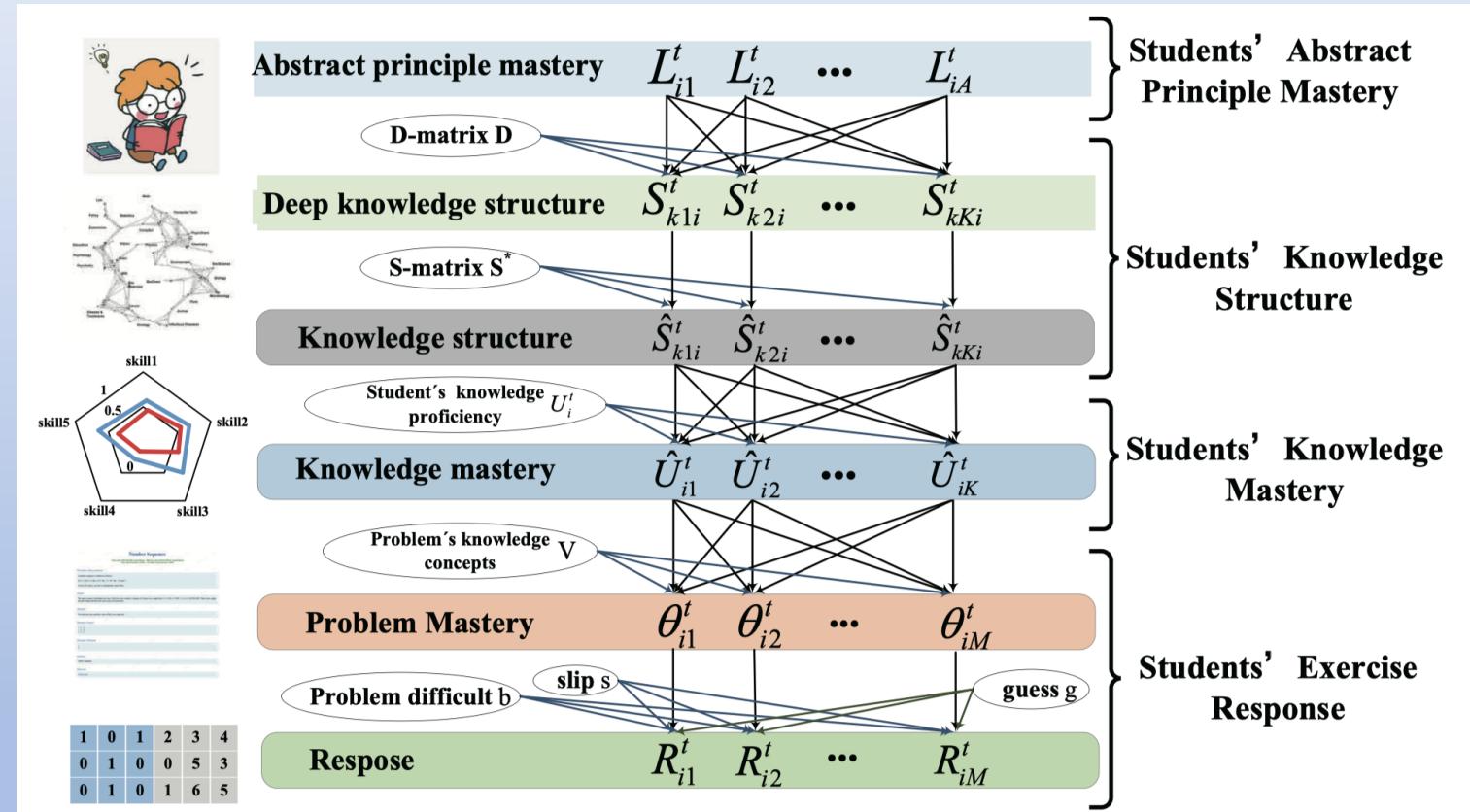
奥苏贝尔
(1918-2008)
美国认知教育心理学家

观察

- **现象 1:** 学习者 u 理解并掌握知识概念 a 和 b ,那么 U 具有很高的概率去掌握未学习但是与 a 和 b 高度相关的知识概念 c 。
- **现象2:** 学习者 u 对题目 e 包含知识概念的熟练度越高，那么 u 对于 e 的掌握程度越高。
- **现象3:** 学习者 u 对于题目 e 的熟练度越高并且题目难度越低，那么 u 答对 e 的概率越高。
- **现象4:** 学习者 u 在真实环境中的练习行为受到猜测 g 与失误因素 s 的影响。

基于学习迁移的知识结构追踪框架

- 学习者抽象原理熟练度
- 学习者知识结构
 - 深度知识结构 (知识概念之间的深度相似性)
 - S-matrix (知识概念之间的文本相似性)
- 学习者知识概念熟练度
 - 知识概念熟练度
 - 学习者的知识结构
- 学习者题目熟练度
 - 题目包含了哪些知识概念
 - 学习者知识概念熟练度
- 学习者练习反馈
 - 学习者题目熟练度
 - 猜测与失误因素



概率图模型

L_{ia} 表示学生 i 对于抽象原理 a 的掌握程度

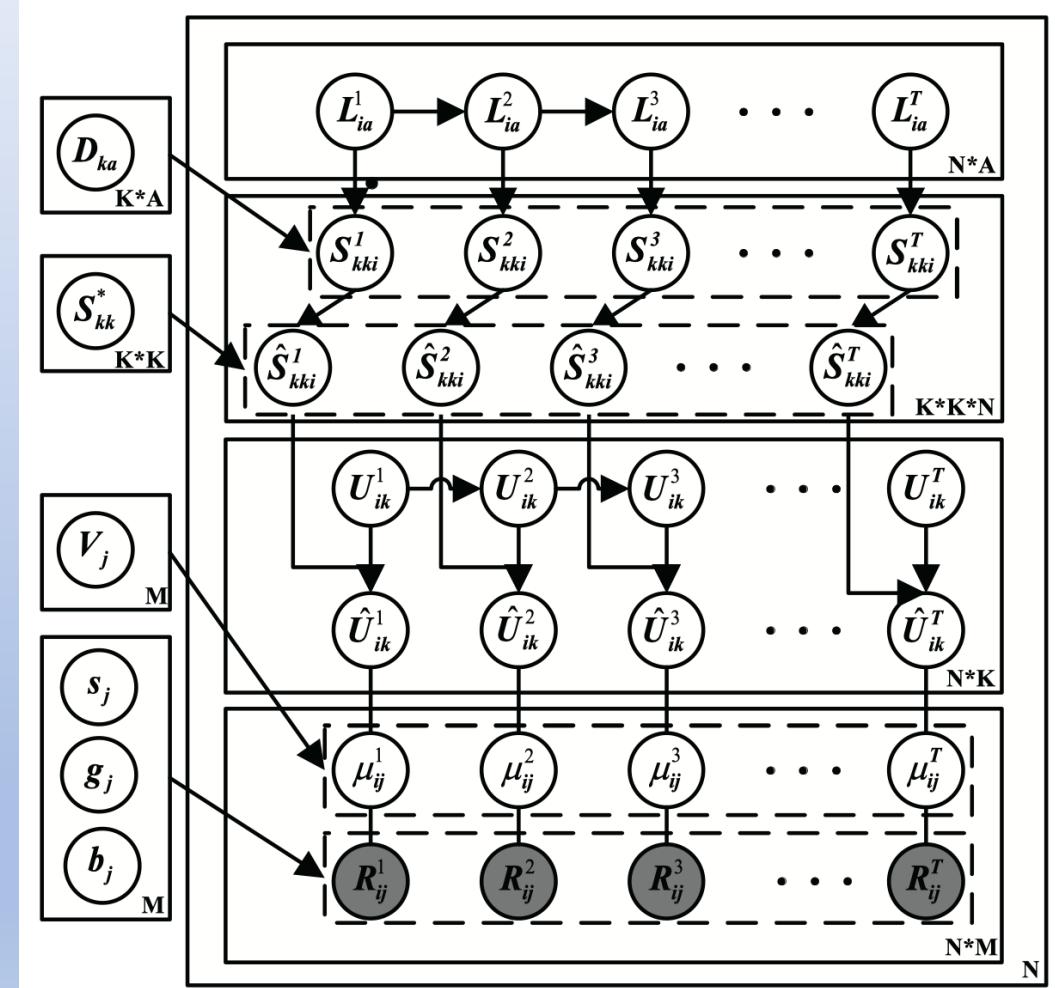
S 表示学生的深度知识结构

S^* 表示知识概念直接的文本相似性

U_{ik} 表示学习者 i 对于 知识概念 k 的熟练度

μ_{ij} 表示学习者 i 对于题目 j 的应对能力

R_{ij} 表示学习者在真实练习反馈中能否答对题目 j , 已经考虑到猜测、失误以及题目难度的结果



模型定义

- 建模学习者练习反馈

$$p(R | \mu, s, g) = \prod_{t=1}^T \prod_{i=1}^N \prod_{j=1}^M [N(R_{ij}^t | (1-s_j)\mu_{ij}^t + g_j(1-\mu_{ij}^t), \sigma_R^2)]^{I_{ij}^t}$$

- 基于IRT理论建模学习者解题能力

$$\mu_{ij}^t = \frac{1}{1 + \exp[-1.7a_j(\theta_{ij}^t - b_j)]}$$

- 基于学习迁移理论建模学习者题目掌握程度

$$\theta^t = \widehat{U}^t * V^T \text{ and } U_{ij}^t = U_{ij}^t + \sum_{k \neq j} \sigma(U_{ik}^t) * S_{kji}^t$$

- 建模学习者知识结构

$$\begin{aligned} S_{ijk}^t &= S_{ijk}^t * \beta_k + S_{ij}^* * (1 - \beta_k) \text{ and } S_{ijk}^t = \frac{\sum_{a=1}^A D_{ia} * D_{ja} * \sigma(L_{ka}^t)^2}{\sqrt{\sum_{a=1}^A (D_{ia} * \sigma(L_{ka}^t))^2} \sqrt{\sum_{a=1}^A (D_{ja} * \sigma(L_{ka}^t))^2}} \end{aligned}$$

模型定义

- 基于学习与遗忘理论建模学习者知识熟练度与抽象原理掌握程度

$$p(U | \sigma_U^2, \sigma_{U_1}^2) = \prod_{i=1}^N N(U_i^1 | 0, \sigma_{U_1}^2 I) \prod_{t=2}^T N(U_i^t | \bar{U}_i^t, \sigma_U^2 I)$$

$$U_{ik}^t = \alpha_i l^t(*) + (1 - \alpha_i) f^t(*), s.t. 0 \leq \alpha_i \leq 1$$

$$p(L | \sigma_L^2, \sigma_{L_1}^2) = \prod_{i=1}^N N(L_i^1 | 0, \sigma_{L_1}^2 I) \prod_{t=2}^T N(L_i^t | L_i^{t-1}, \sigma_L^2 I)$$

- 建模学习与遗忘过程

$$l^t(*) = U_{ik}^{t-1} \frac{G * f_k^t}{f_k^t + r} \text{ and } f^t(*) = U_{ik}^{t-1} e^{-\frac{\Delta t}{P}}$$

- 以Q-matrix 作为先验建模题目知识概念矩阵

$$\ln p(V | Q) = \ln \prod_{(j,q,p)} p(>_j^+ | V) p(V) = \prod_{j=1}^M \prod_{q=1}^K \prod_{p=1}^K I(q >_j^+ p) \ln \frac{1}{1 + e^{-1(V_{jq} - V_{jp})}} - \frac{1}{2\sigma_V^2} |V|_F^2$$

- 其中 $>_j^+$ 定义为 $q >_j^+ p, if Q_{jq} = 1 and Q_{jp} = 0$

模型训练

- 最大化模型后验概率 $p(U, L, D, V, S^*, \alpha, \beta, a, b, s, g | R, Q)$

$$\begin{aligned} & p(U, L, D, V, S^*, \alpha, \beta, a, b, s, g | R, Q) \\ & \propto p(R | \hat{R}) \times p(V | Q) \times p(U | \sigma_U^2, \sigma_{U1}^2) \times p(L | \sigma_L^2, \sigma_{L1}^2) \times p(D | \sigma_D) \times p(S^* | \sigma_S) \end{aligned}$$

- 其中 \hat{R} 定义如下

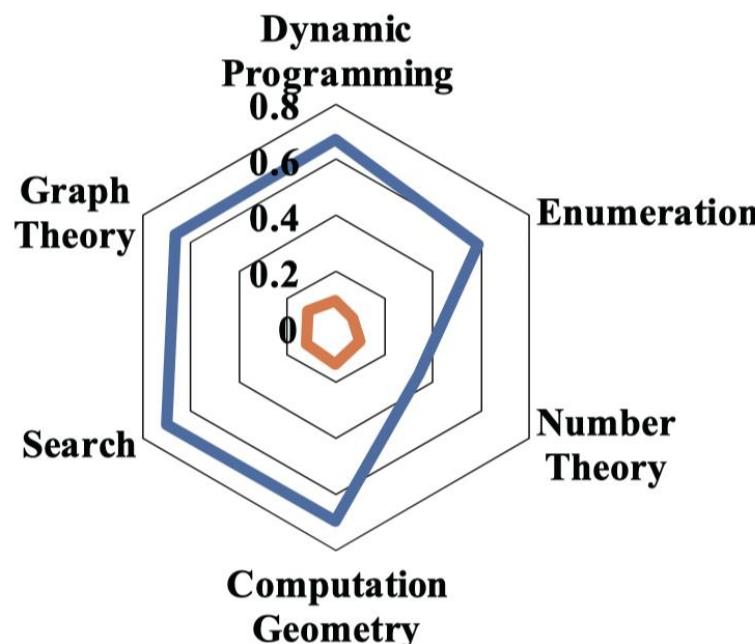
$$\hat{R}_{ij}^t = (1 - s_i - g_j) \frac{1}{1 + \exp[-1.7 a_j (\hat{U}_i^t V_j - b_j)]} + g_j$$

- 最终优化目标

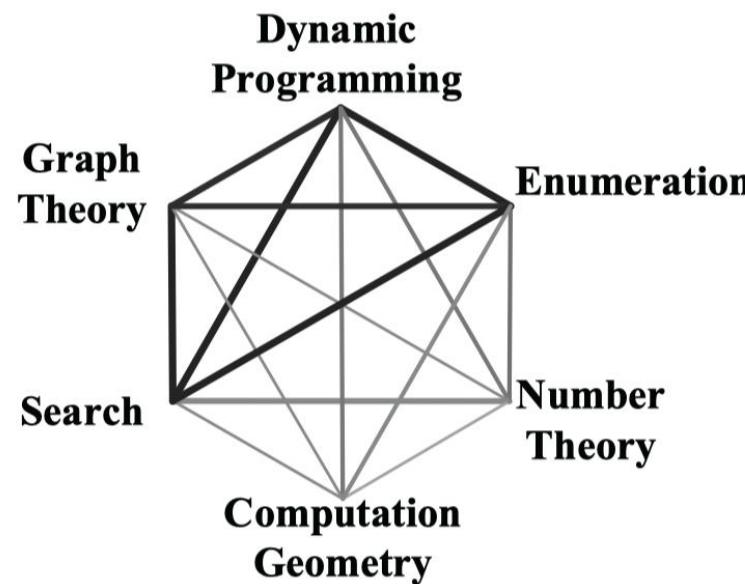
$$\begin{aligned} \min_{\Phi} \mathcal{E}(\Phi) = & \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^M I_{ij}^t \left[\hat{R}_{ij}^t - R_{ij}^t \right]^2 \\ & - \lambda_P \sum_{j=1}^M \sum_{q=1}^K \sum_{p=1}^K I(q >_j^+ p) \ln \frac{1}{1 + e^{-(V_{jq} - V_{jp})}} \\ & + \frac{\lambda_U}{2} \sum_{t=2}^T \sum_{i=1}^N \left\| \overline{U}_i^t - U_i^t \right\|_F^2 + \frac{\lambda_{U1}}{2} \sum_{i=1}^N \|U_i^1\|_F^2 \\ & + \frac{\lambda_L}{2} \sum_{t=2}^T \sum_{i=1}^N \left\| L_i^{t-1} - L_i^t \right\|_F^2 + \frac{\lambda_{L1}}{2} \sum_{i=1}^N \|L_i^1\|_F^2 \\ & + \frac{\lambda_V}{2} \sum_{i=1}^M \|V_i\|_F^2 + \frac{\lambda_S}{2} \sum_{i=1}^K \|S_i^*\|_F^2 + \frac{\lambda_D}{2} \sum_{i=1}^K \|D_i\|_F^2 \end{aligned}$$

诊断结构实例

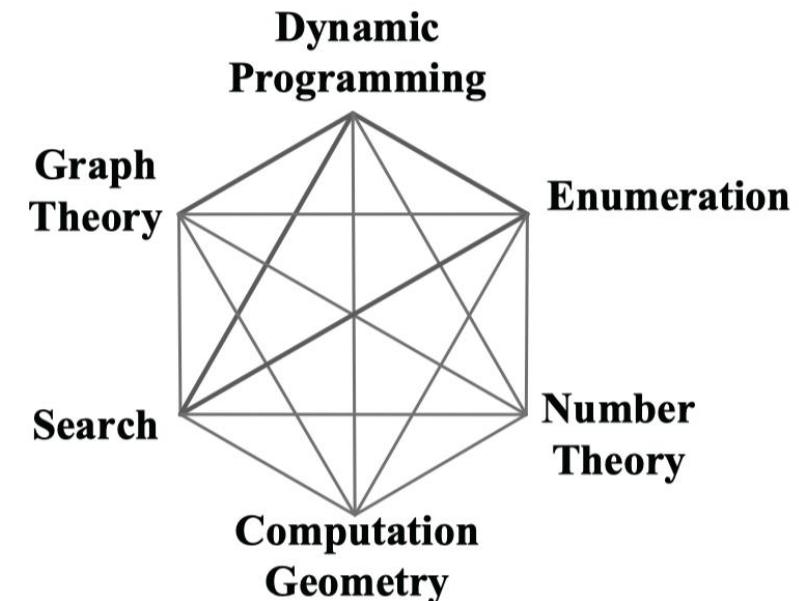
—Student A — Student B



A B 两个学习者对于6种
知识概念的熟练度

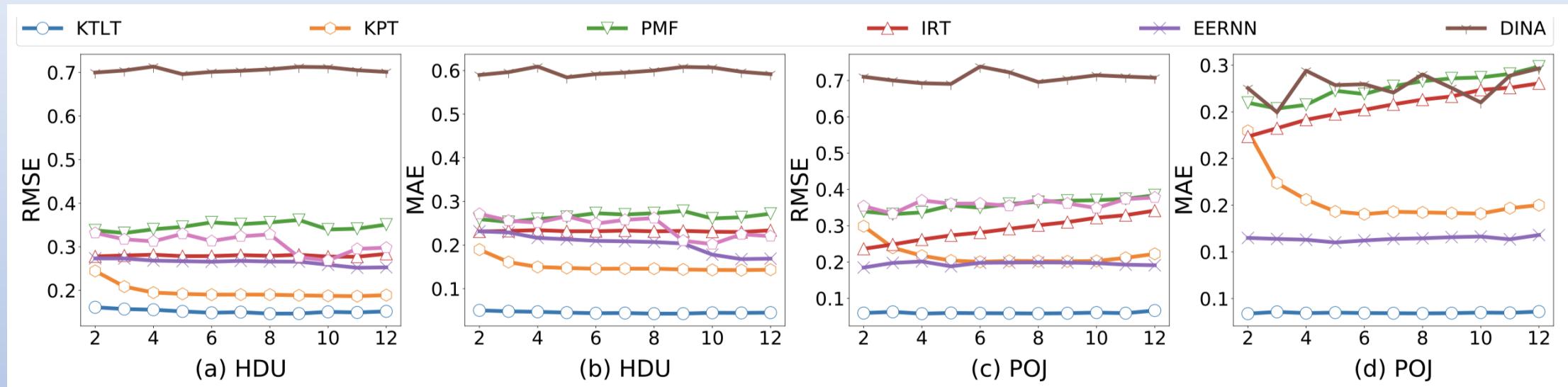


学习者A的知识结构



学习者B的知识结构

不同知识追踪模型之间的对比



- 对比模型：KPT、PMF、IRT、EERNN、DINA
- 数据集：HDU（杭电OJ平台）、POJ（北大OJ平台）

数据集	#学生	#知识概念	#习题	#提交记录	#人平均记录
HDU	5714	77	2434	366,836	64
POJ	2180	10	1128	158,885	72

总结

- 实验结果：
 - 可有效追踪学习者的知识结构、抽象原理掌握程度与知识熟练度
 - 可有效地捕捉学习者的学习迁移情况，从而较为准确地预测学生的成绩
- 今后工作：
 - 实时追踪学习者的知识掌握水平->建立基于隐马尔可夫模型的概率模型，提高实时性
 - 解释学习者的抽象原理掌握程度->采用教育心理学理论提炼抽象原理，增加可解释性

敬请批评指正

谢谢