

# Knowledge Tracing

Jiaming Zhang

# Outline

- Introduction: Knowledge Tracing
- Bayesian Knowledge Tracing(BKT)
- Deep Knowledge Tracing(DKT)
- Improvements based on DKT
- Association with real application scenery

# Introduction: Knowledge Tracing

- 知识追踪：基于学生行为序列  $(x_0, \dots, x_t)$  进行建模，预测学生对知识的掌握程度  $x_{t+1}$ ，这里  $\mathbf{x}_t = \{q_t, a_t\}$ 
  - 知识点（考点）  $q_t$ ：言语、数量、判断、常识、资料
  - 行为标签  $a_t$ ：代表题目回答是否正确，取值范围为  $\{0,1\}$
  - 学生行为序列  $(x_0, \dots, x_t)$ ：粗粒度可按练习划分，细粒度可按题目划分
- 知识追踪的应用场景
  - 学习路径规划
  - 预测分计算
  - 知识图谱构建

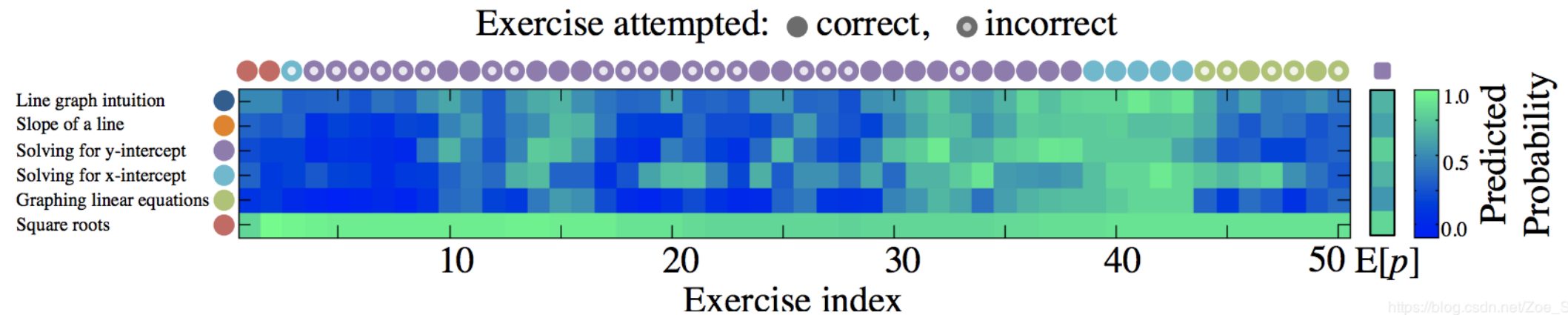
# Bayesian Knowledge Tracing

- Knowledge tracing: Modeling the acquisition of procedural knowledge: 1995 User Modeling and User-Adapted Interaction
- 基本假设：用户掌握了某个知识点之后，之后不会遗忘，即标识该知识点的隐变量的取值由0变成1之后就保持不变了
- 采用一阶隐马尔可夫模型（HMM）求解，根据状态转移矩阵来预测下一个状态，根据当前的状态来预测学生的答题结果
- BKT的主要缺陷：
  - BKT只对于单个知识点进行建模，无法表征多个知识点交叉重复练习的现实场景
  - 基本假设用户不遗忘知识点，符合用户的心理预期，但不符合实际情况
  - 无法表征知识点之间的内在相关性

# Deep Knowledge Tracing(1)

- Deep Knowledge Tracing: 2015 NeurIPS
- 循环神经网络（RNN）是一种时间序列的模型，能够利用早期的信息进行预测，天然具有高维连续的隐状态表示的特征
- 在2015年Chris Piech等人将RNN应用于知识追踪领域，提出了基于RNN的DKT算法

# Deep Knowledge Tracing(1)



- 上图描述了一个学生在学习八年级数学课室的知识追踪可视化展示
- 这个学生一开始答对了两个关于平方根的问题，然后答错了一道关于X-轴截距的题，接下来做了一系列关于X-轴截距、Y-轴截距和线性方程作图的题目（在这里平方根、X-轴截距、Y-轴截距等都可以看作是一个知识点）
- 在学生做完每一道练习时我们都可以预测他在下一道题目上的表现，题目可以是来源于不同知识点的题目。在这个图中我们只预测了和当前练习题相关的知识点的掌握情况

# Deep Knowledge Tracing(1)

- 作者采用了传统的RNN模型和其变种LSTM，其输入数据为经过编码的题目序列 $q_t$ ，输出为 $a_t$ ，对于输入的编码方式有以下两种：
  - 将输入进行one-hot编码，如模型输入数据涉及 $M$ 个知识点，每道题有两种结果0, 1（分别对应答错和答对），则模型输入长度为 $2M$ 
    - e.g. 对于某题，其知识点为 $i$ ，若答对，对应输入为第 $M+i$ 位为1其余位置为0；若答错，则第 $i$ 位为1其余位置为0
  - 若知识点维度 $M$ 特别大时，则将输入数据进行压缩，可采用压缩感知使输入数据维度从 $2M$ 降低至 $\log(2M)$

# Deep Knowledge Tracing(1)

- DKT的优点
  - 能够反应长时间的知识关系，基于RNN的特性能够根据学生近期学习表现进行预测（近因效应），也能根据实际学生学习路径进行建模
  - 能够对复杂的知识点之间的联系进行建模，如构建知识图谱
  - 能够处理多知识点的问题
- DKT的缺点
  - 模型无法重构输入，即可能出现输入某一知识点答题错误，模型对该知识成分的预测反而是正确
  - 在时间序列上，学生对知识点的掌握程度不具有连续一致性，波动情况较大



# Deep Knowledge Tracing(2)

- Going Deeper with Deep Knowledge Tracing:
  - 2016年Xiaolu Xiong等人对DKT和PFA(Performance Factor Analysis), BKT模型进行了比较, 对DKT模型能碾压其他两种模型的结果进行了怀疑并加以详细的实验论证, 进一步讨论了原论文能够得出上述结果的原因, 对进一步使用DKT模型提供了参考
  - DKT文中实验部分所用的ASSISTments数据集中存在23.6%的重复数据, 这部分数据应该舍弃而不应该用于训练或测试
  - DKT模型在实验中, 并没有去除脚手架式的教学问题的做题记录, 这就导致DKT模型能够有更多信息引入模型
  - 由于DKT处理多知识成分的问题时, 单条做题记录会被扩展成多条, 存在重复利用数据的问题
  - 结论是: DKT相比BKT和PFA没有碾压式的超越, 但是的确会比其他模型结果要好

# Deep Knowledge Tracing(3)

- Addressing Two Problems in Deep Knowledge Tracing via Prediction-Consistent Regularization: 2018 International Conference on Education Data Mining
  - 2018年Chun-Kit Yeung等人对DKT的两种缺陷，即对输入序列存在重构问题和预测结果的波动性问题，用增加相应正则项的方式进行改进，提出了DKT+算法
  - 针对重构问题，引入时间 $t$ 时刻的输入值正则项（DKT只考虑 $t$ 时刻输出和 $t+1$ 时刻输入）
  - 针对波动性问题，引入L1和L2正则，平滑输出预测结果
  - 由于DKT处理多知识成分的问题时，单条做题记录会被扩展成多条，存在重复利用数据的问题

# Deep Knowledge Tracing(4)

- Prerequisite-Driven Deep Knowledge Tracing: 2018 ICDM
  - 2018年Penghe Chen等人在使用DKT时考虑到数据的稀疏性(知识点空间比较大, 学生做题比较有限), 为解决由于数据稀疏性带来的模型评估不准确, 提出了将知识结构的信总纳入模型来解决上述问题
  - 基本思路: 如果 $K_1$ 是 $K_2$ 的前置知识点, 则 $K_2$ 的掌握程度要小于等于 $K_1$ , 即后置的掌握程度要小于等于前置的掌握程度
  - 具体过程描述: 学生 $i$ 如果在 $t+1$ 时刻掌握了 $K_2$ , 则说明在前一时刻 $t$ , 学生已经掌握了其前置知识点 $K_1$ ; 反之, 若学生 $i$ 在 $t$ 时刻没有掌握 $K_1$ , 则说明在下一时刻 $t+1$ , 其更不可能掌握后置知识点 $K_2$

# Association with real application scenery

- 目前题库所用预测分算法基本模型IRT，其基本假设与BKT一致，且提出年代更早(1990)，求解方式更简单(MLE)
- DKT虽然性能要比BKT&IRT更好，但其存在的两大缺陷使其无法直接应用于已产品化的预测分计算
  - 注：2016年Kevin H. Wilson等人发表的文章指出在某些数据集上，改进版的IRT方法是能比DKT取得更好的实验结果的
- 在其他应用场景上DKT会有用武之地？ e.g. 特训营学习路径规划

**Thank you !**  
**Any questions?**