Exploiting Cognitive Structure for Adaptive Learning

Liu, Qi, Shiwei Tong, Chuanren Liu, Hongke Zhao, Enhong Chen, Haiping Ma, and Shijin Wang. "Exploiting cognitive structure for adaptive learning." In *Proceedings of the 25th ACM **SIGKDD** International Conference on Knowledge Discovery & Data Mining*, pp. 627-635. **2019.**

论文连接：https://dl.acm.org/doi/abs/10.1145/3292500.3330922

目录

背景：自适应学习，又称自适应教学，依赖于学习路径推荐，它依次推荐个性化的学习项目(如讲座、练习)，以满足每个学习者的独特需求。

（一）目的：根据① **学习者的知识水平**和② **知识结构** 进行 学习路径推荐；

（二）问题定义与符号约定；

（三）模型：（1）知识追踪：用于追踪学习者知识状态变化的过程；

（2）认知导航（Cognitive Navigation）根据知识结构快速选择潜在候选项，起到作用：

① 防止推荐的学习项目违背路径的逻辑性；

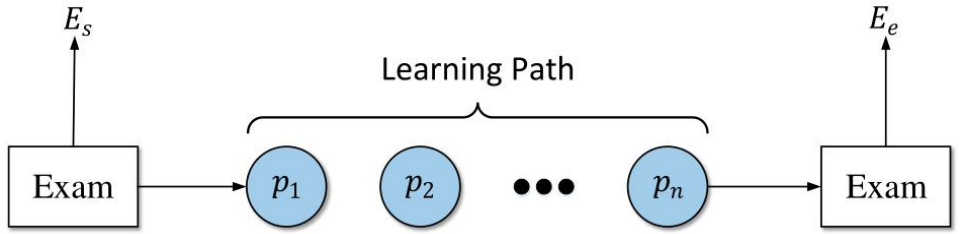
② 减少在探索过程中的搜素空间。

（3）Actor-Critic推荐：根据学生目前的知识水平从候选项中做选择。

（四）实验：（1）**数据集/模拟器的构建**；

（2）实验结果。

问题定义与符号约定



Learning Session：学习是一个漫长的过程，由若干个学习环节(learning session)和某个学习目标组成，每个学习环节如图定义，它包含两个主要部分
（1）Exam：检索学习路径的学习效果；
（2）Learning Path：由若干个学习项目p_i组成

定义一个学习session的***学习效果***如下：

$$E_{\mathcal{P}} = \frac{E_e - E_s}{E_{sup} - E_s},$$

其中E_{sup}是考试的满分。

$\mathcal{H} = \{h_0, h_1, ..., h_m\}$ ：在以往learning session中生成的历史学习记录；

$h = \{k, score\}$

$\mathcal{P} = \{p_0, p_1, ...p_N\}$ ：学习路径；

$\mathcal{F}_i = \{p_i, score_i\}$ ：第i步的推荐内容及学生回答得分。

Learning Path Recommendation Problem：给定历史学习记录$\mathcal{H}$，学习者的一个确定的学习目标$\mathcal{T}$，先序关系图G，我们的目标是推荐一条长度为N的学习路径，使得整个学习路径的效能$E_{\mathcal{P}}$最大化。

模型CSEAL

Overview

State：第i步的状态state_i定义为目标$\mathcal{T}$和知识水平S_i的结合。 $state_i = \mathcal{S}_i \oplus \mathcal{T}$

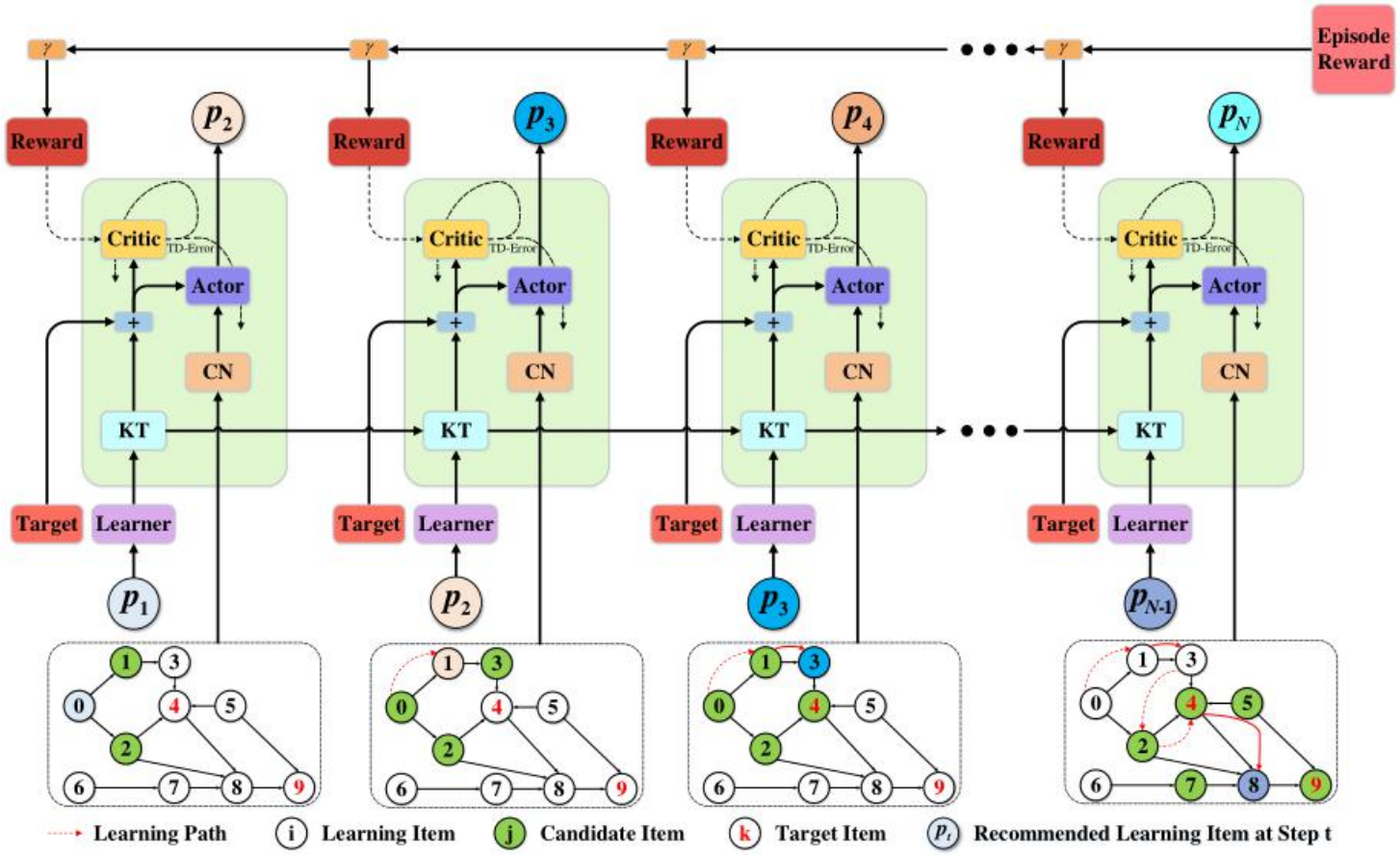
$\mathcal{T} = \{0,1\}^M$，M为知识点的个数。

Action：CSEAL在第i步采取的行动a_i是随机抽样而来的，服从分布$\pi(a|state_i;\theta) = P(a|\mathcal{H},\mathcal{F}_{0,\dots,i-1},\mathcal{T};\theta)$，其中$\theta$为模型参数集。

Reward：定义reward$R_i = \sum_{j=i}^{N} \gamma^j r_j$

γ：折扣因子；r_j单步奖励

模型CSEAL



（1）知识追踪：用于追踪学习者知识状态变化的过程；

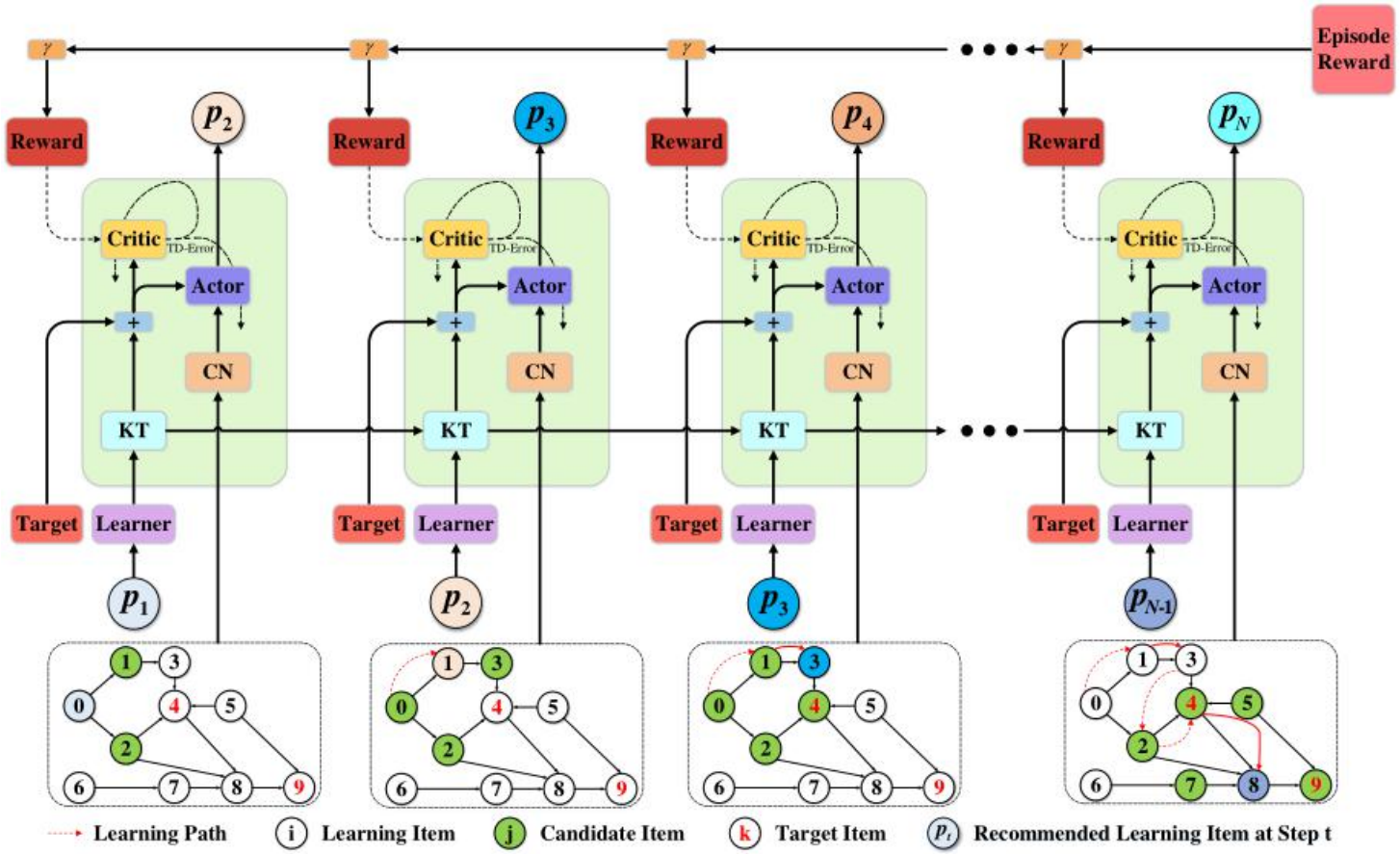
（2）认知导航（Cognitive Navigation）根据知识结构快速选择潜在候选项，起到作用：

① 防止推荐的学习项目违背路径的逻辑性；

② 减少在探索过程中的搜素空间。

（3）Actor-Critic推荐：根据学生目前的知识水平从候选项中做选择。

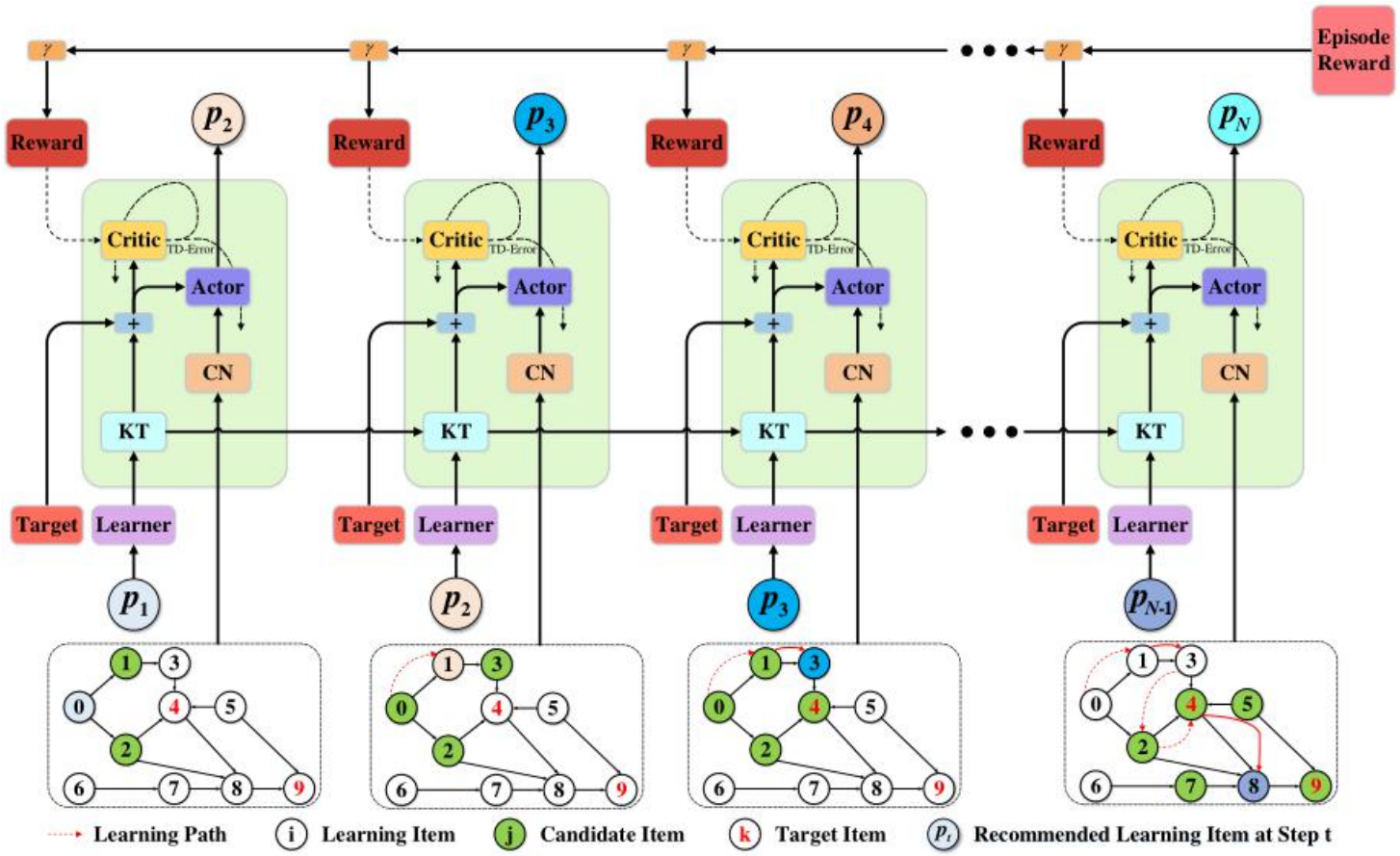
模型CSEAL



（1）知识追踪：用于追踪学习者知识状态变化的过程；

S_t：t时刻知识状态

模型CSEAL



认知导航（Cognitive Navigation）根据知识结构快速选择潜在候选项，起到作用：

① 防止推荐的学习项目违背路径的逻辑性；

② 减少在探索过程中的搜素空间。

选取$k = 2$

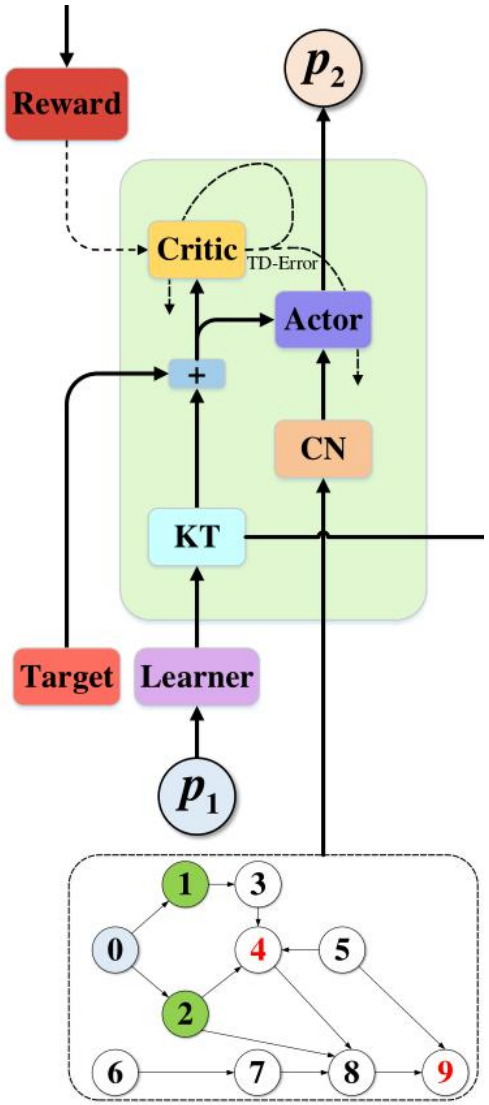
Algorithm 1 Cognitive Navigation.

Input: central focus C, a prerequisite graph G, learning target $\mathcal{T}$
Output: $\forall d \in D$ has a path to $\mathcal{T}$ in G and is one of the k-hop neighbors of C

Initialize candidates $\mathcal{D} = \varnothing$, $Q = \varnothing$;
add C to $\mathcal{D}$;
add successors within $k-1$ hop of C to $\mathcal{D}$;
add predecessors within $k-1$ hop of C to Q;
while $Q \neq \varnothing$ **do**
  $q \leftarrow Q$.pop();
  add q to D;
  add neighbors of q to D;
end while
for d in D **do**
  if d can not reach T **then**
    del d from D;
  end if
end for
return $\mathcal{D}$

第一步的central focus $\mathcal{C}$：$\mathcal{H}$的最后一个/手动指定/没有前驱结点的点。

模型CSEAL

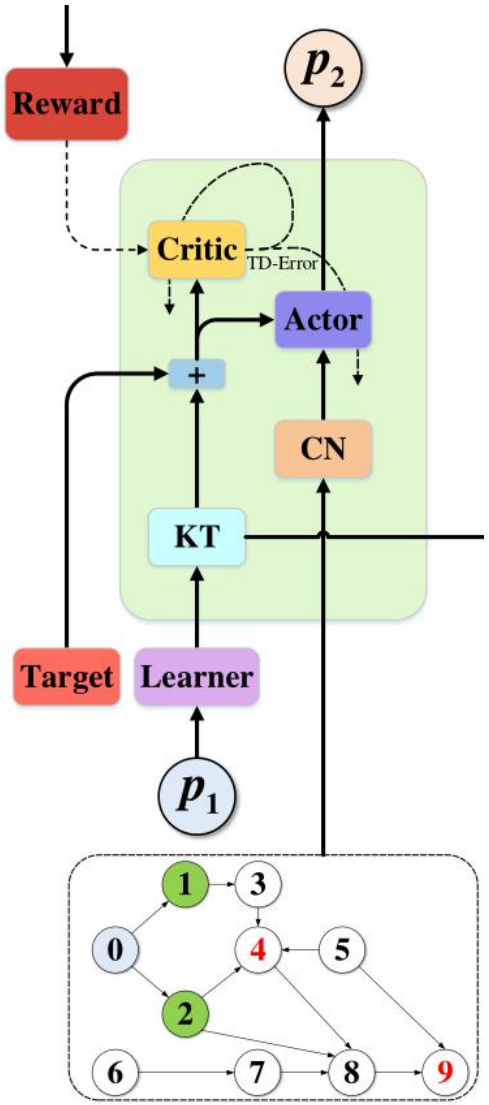


Actor-Critic推荐：根据学生目前的知识水平从候选项中做选择。

Actor和Critic是两个不同的神经网络。
Actor：用于预测行为的概率(Policy Gradient);
Critic：预测在这个状态下的价值(value-based) $\mathcal{V}(\cdot;\theta_{\mathcal{V}})$.

过程：
① Actor基于概率在CN提供的候选集中选行为（策略函数$\pi(a|state;\theta)$）
② Critic估计每一个状态（$state$）的价值，用这个状态的价值减去下个状态的价值（TD-error）。
③ TD-error如果是正的，下个动作就要加大更新，如果是负的，就减小actor的更新幅度。
Criric基于Actor的行为评判行为的得分，Actor根据Critic的评分修改选行为的概率。
总结：
演员(Actor)是指策略函数$\pi_\theta(a|state)$，即学习一个策略来得到尽量高的回报。（2）评论家(Critic)是指值函数$V_\pi(state)$，对当前策略的值函数进行估计，即评估Actor的好坏。

模型CSEAL



Actor-Critic推荐：根据学生目前的知识水平从候选项中做选择。

通过步骤i估计的预期收益：

$$v_i = \mathcal{V}(state_i; \theta_v)$$

使用第i步的policy gradient更新参数

$$\nabla\theta = log\pi(a|state;\theta)(R_i - v_i)$$

通过优化实际收益与估计值之间的差来训练value network：

$$\mathcal{L}oss_{value} = \|v_i - R_i\|_2^2.$$

为了保证收敛，设置损失函数：

$$\mathcal{L}oss = \|\mathcal{V}(state_i;\theta_v) - R_i\|_2^2 + \alpha \cdot -log\ \pi(a|state_i;\theta)(R_i - v_i) + \beta \cdot -log\ \pi(a|state_i;\theta)R_i,$$

数据集/模拟器的构建

选用数据集：Junyi数据集；本论文中仿照DKT合成的数据集

Junyi数据集

Table 1: The statistics of the dataset.

Statistics	Value
number of learners	247,548
number of sessions	525,062
number of learner logs	39,462,202
number of exercises in learner logs correctly answered	21,460,360
median of exercises in one session	21
median of knowledge concepts in one session	3
median of practice frequency on a concept in one session	8
number of nodes in graph	835
number of links in graph	978

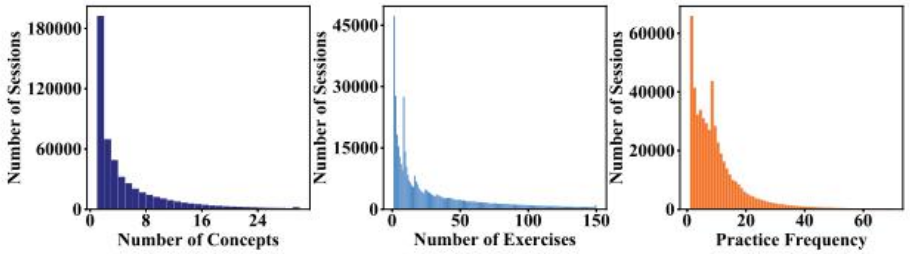


Figure 5: Distributions of Sessions.

合成数据集

构造10个知识点，12条边，学习者的能力θ值以及演化规则（由教研专家设计得出），学习目标从点中随机取出

$$P(\theta) = c + \frac{1-c}{1+e^{-Da(\theta-b)}},$$

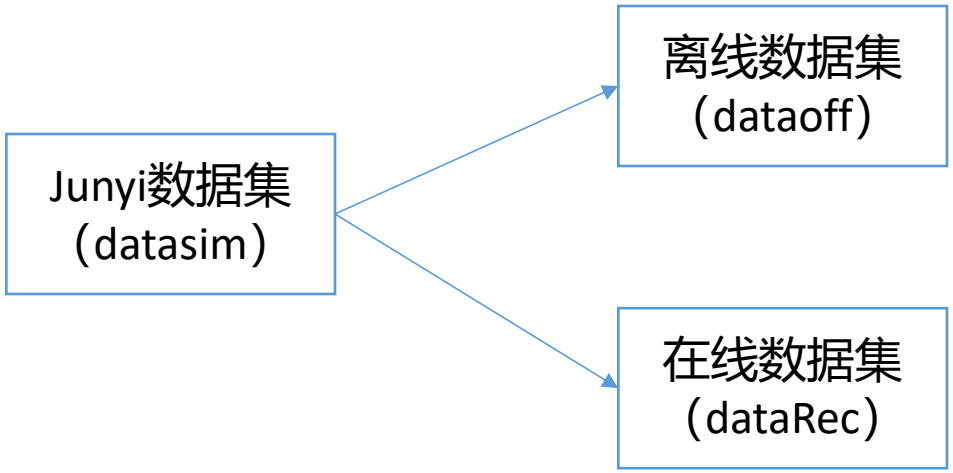
根据3参数的IRT公式，构造max length 为50的4000条记录。

NOTE：需要注意的是，由于学生回答的正确性及各类参数完全由IRT公式给出，实际上也构建了一个**模拟器(KSS)**，可以用于分析某个练习序列中**未包含的练习**是否能够正确回答，直接用作评估强化学习模型中的学习路径（例如计算提升）或训练代理（例如CSEAL和其他基线）的环境。

80/10/10: train/val/test

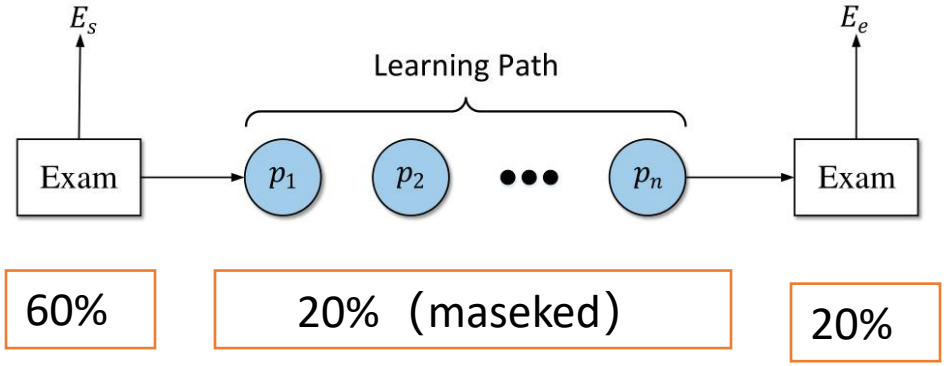
数据集/模拟器的构建

第二种模拟器的构建方式（Knowledge Evolution based Simulator, KES）：使用DKT模型的预测值。



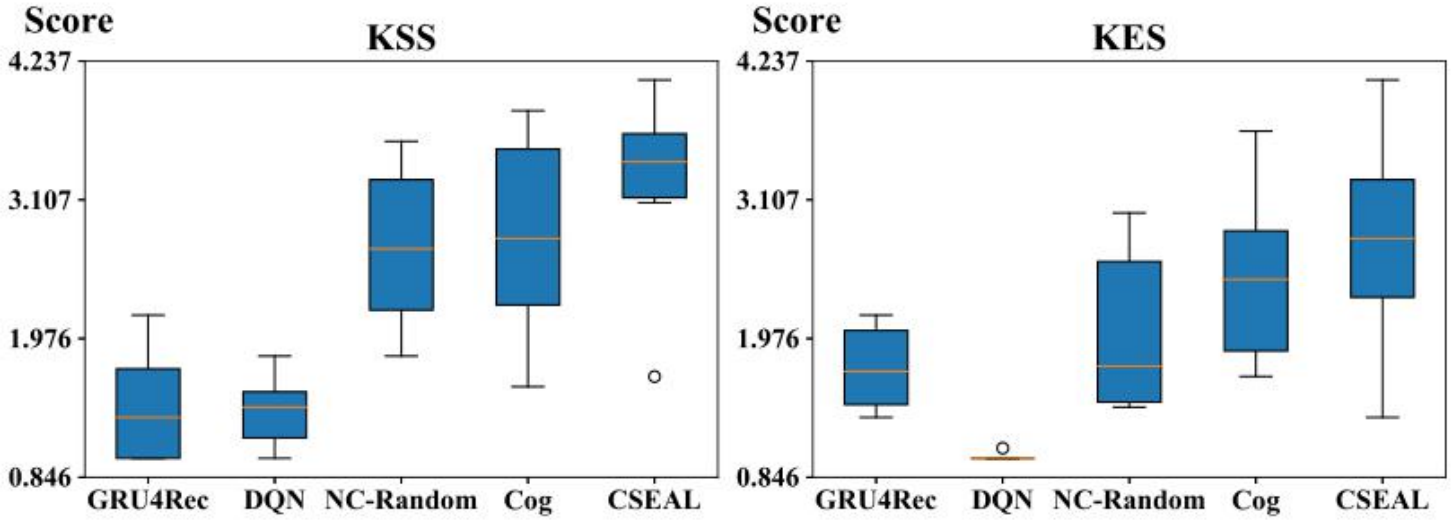
Randomly 50%/50%

用于训练基线模型和Knowledge Tracing模型



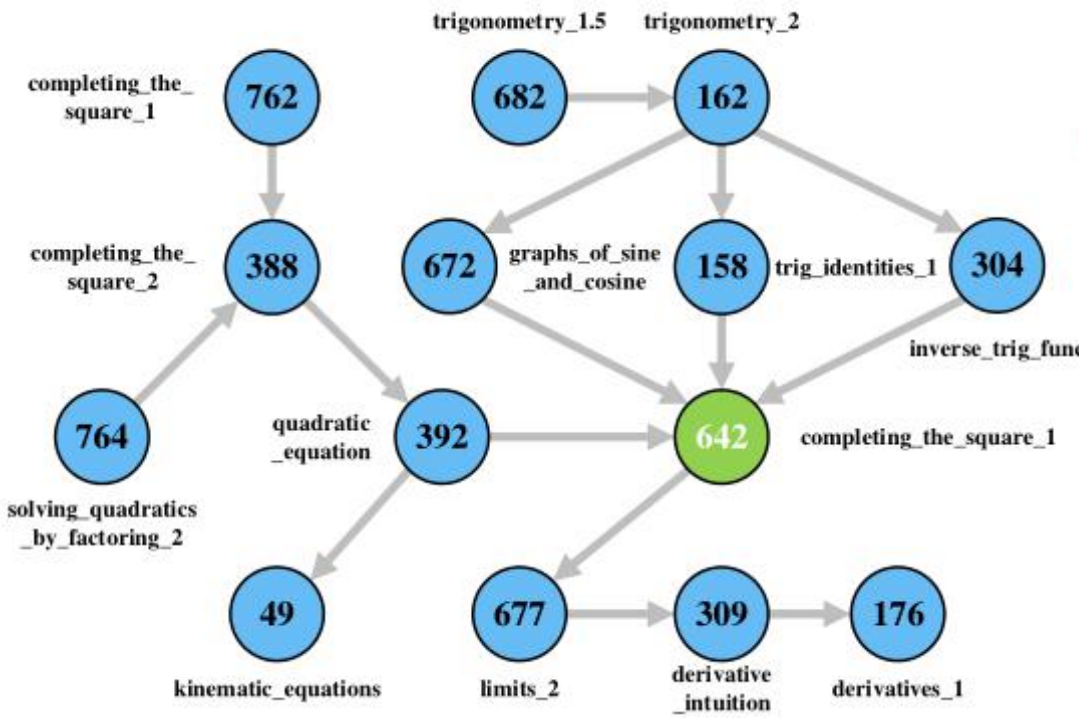
上述所有数据，80/10/10，train/val/test

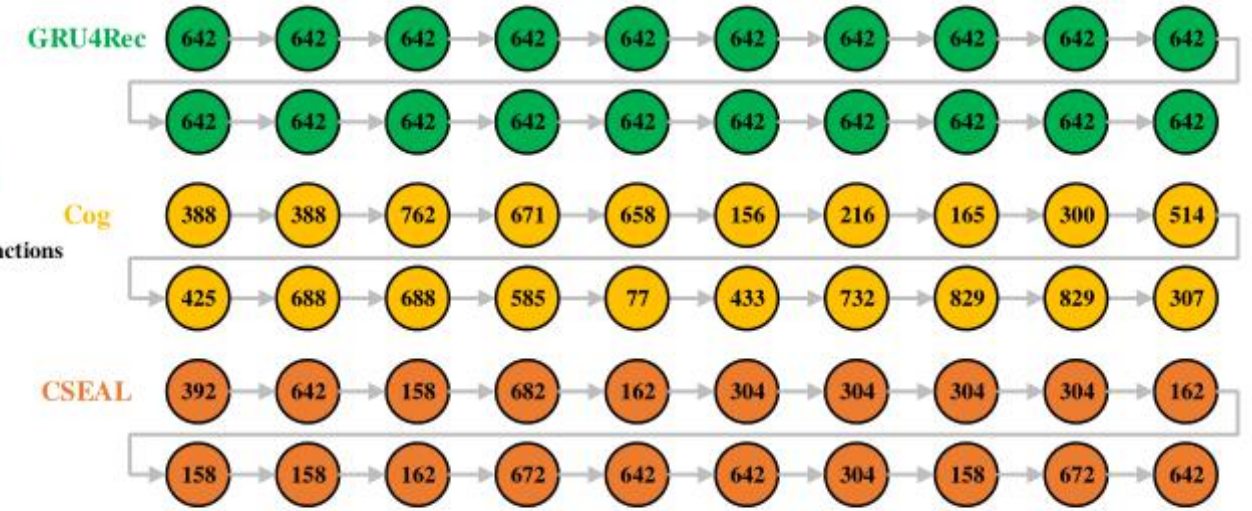
实验结果



GRU4Rec：一种经典的基于会话的推荐模型。模型的输入是会话的顺序，而输出是下一步出现的学习项目的概率分布；
DQN：一些工作已经提出利用强化学习来解决这个问题，但是这些工作需要丰富的人类领域知识来设计MDP中的转移矩阵和精确的初始状态，这是不实际的。因此，KT模型和深度Q学习分别取代了MDP和简单Q学习（value-based 中的单步reward）中的状态。
CN Random：推荐项目是从CN选择的候选项目中随机选取的。
Cog：从CN选择的候选项目中随机选取推荐项目进行加权，其中项目的权重与KT模型测量的掌握程度成反比（掌握的不好，权重高）

实验结果


trigonometry_1.5
trigonometry_2
completing_the_square_1
762
682
162
completing_the_square_2
388
672
graphs_of_sine_and_cosine
158
trig_identities_1
304
inverse_trig_functions
764
solving_quadratics_by_factoring_2
quadratic_equation
392
642
completing_the_square_1
49
kinematic_equations
677
limits_2
309
derivative_intuition
176
derivatives_1


GRU4Rec
642 642 642 642 642 642 642 642 642 642
642 642 642 642 642 642 642 642 642 642
Cog
388 388 762 671 658 156 216 165 300 514
425 688 688 585 77 433 732 829 829 307
CSEAL
392 642 158 682 162 304 304 304 304 162
158 158 162 672 642 642 304 158 672 642

谢谢！