

FKA-DKT: 融合知识与能力的深度知识追踪模型

陈 成¹, 董永权^{1,2,3}, 贾 瑞¹, 刘 源¹

(1. 江苏师范大学计算机科学与技术学院, 江苏 徐州 221116)

(2. 江苏省教育信息化工程技术研究中心, 江苏 徐州 221116)

(3. 徐州市云计算工程技术研究中心, 江苏 徐州 221116)

[摘要] 知识追踪(KT)是智能教育中的一个重要研究问题,其通过分析学生的历史交互来预测其未来的答题情况. 现有的主流KT模型仅根据学生的知识掌握情况对学生进行建模,忽视了学生的个人能力在答题中的作用. 因此,本文提出了一种融合知识和能力的深度知识追踪模型(FKA-DKT). 首先利用DKT模型构建基于知识的答题预测网络(KAPN),从知识层面预测学生答案的正确性. 然后提出基于能力的答案预测(AAPN)网络对学生的能力进行建模,从能力层面预测学生答案的正确性. 最后,将KAPN和AAPN的预测结果进行线性组合,使模型能够融合知识和能力两个方面的信息来预测学生的作答结果. 在4个公开的数据集上的实验结果表明,相较于现有的主流方法,FKA-DKT在AUC指标上取得了显著的性能提升.

[关键词] 知识追踪,深度知识追踪,个人能力建模

[中图分类号]TP391 [文献标志码]A [文章编号]1001-4616(2024)02-0129-11

FKA-DKT: Deep Knowledge Tracing Model Based on the Fusion of Knowledge and Ability

Chen Cheng¹, Dong Yongquan^{1,2,3}, Jia Rui¹, Liu Yuan¹

(1. College of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, China)

(2. Jiangsu Engineering Technology Research Center of ICT in Education, Xuzhou 221116, China)

(3. Xuzhou Cloud Computing Engineering Technology Research Center, Xuzhou 221116, China)

Abstract: Knowledge tracing (KT) is an important research problem in intelligent education, which predicts students' future answering behaviors by analyzing their historical interactions. Existing mainstream KT models only model students based on their knowledge mastery, neglecting the role of students' personal abilities in answering questions. Therefore, this paper proposes a deep knowledge tracing model (FKA-DKT) that integrates both knowledge and ability. First, we use the DKT model to construct a Knowledge-based Answer Prediction Network (KAPN), which predicts student answer correctness at the knowledge level. Then, we propose an Ability-based Answer Prediction Network (AAPN) to model students' abilities and predict answer correctness at the ability level. Finally, we linearly combine the predictions from KAPN and AAPN to integrate both knowledge and ability information for answer prediction. Experimental results on four publicly available datasets show that compared to existing mainstream methods, FKA-DKT achieves significant performance improvements in terms of the AUC metric.

Key words: knowledge tracing, deep knowledge tracing, personal ability modeling

知识追踪(knowledge tracing, KT)^[1]是教育数据挖掘的一个重要的研究方向,其目标主要是通过建立学生知识状态随时间变化的模型,来判断学生对知识的掌握程度,并从学生的学习轨迹中挖掘出潜在的学习规律,从而提供个性化的指导,以达到人工智能辅助教育的目的^[2]. 近年来,随着智能辅导系统(intelligent tutoring system, ITS)和大规模在线开放课程(massive open online courses, MOOCs)等在线教育平

收稿日期:2023-06-01.

基金项目:国家自然科学基金面上项目(61872168)、江苏省教育科学十四五规划项目(d/2021/01/112)、江苏师范大学研究生科研与实践创新计划项目(2022XKT1527).

通讯作者:董永权,博士,教授,研究方向:数据集成、数据挖掘、群体智能、教育信息化等. E-mail:tomdyq@jsnu.edu.cn

台的发展和普及,数百万的用户选择通过在线教育平台进行学习,同时 2020 年爆发的新冠疫情进一步推动了在线教育行业的发展. 在此背景下,在线教育平台得以收集并保存海量的教育相关数据资源,为 KT 的研究提供了可靠的数据来源,进而推动 KT 的发展.

过去在 KT 领域常使用基于隐马尔可夫模型(hidden markov model,HMM)的传统机器学习模型对学生的知识状态进行建模,并取得了一定的效果. 后来深度学习以其强大的特征提取能力受到了研究者的广泛关注,许多研究者开始将深度学习应用到 KT 领域,称为基于深度学习的知识追踪(deep learning based knowledge tracing,DLKT). 相对于传统的机器学习模型,DLKT 能够捕捉到更为复杂的学生知识表征^[3],并取得更好的效果. 目前对 DLKT 的研究已经成为 KT 领域的一大热点.

然而,已有的多数 DLKT 方法^[3-6]只关注学生学习过程中在知识层面的变化情况,仅对学生的知识状态建模,并未注意到学生的个人能力在答题过程中的作用. 事实上,学生的个人能力,如计算能力和阅读理解能力,也是学习过程中不可或缺的因素,而能力较强的学生更容易答对这些能力要求高的题目. 例如,在图 1 中,问题 1 是一道仅涉及计算能力的加法题,而问题 2 是一道考察阅读理解能力的减法应用题. 阅读理解能力较差的学生即使掌握了减法的知识,但会因为无法理解题干含义而答错问题 2. 因此,针对学生个人能力的建模显得至关重要.

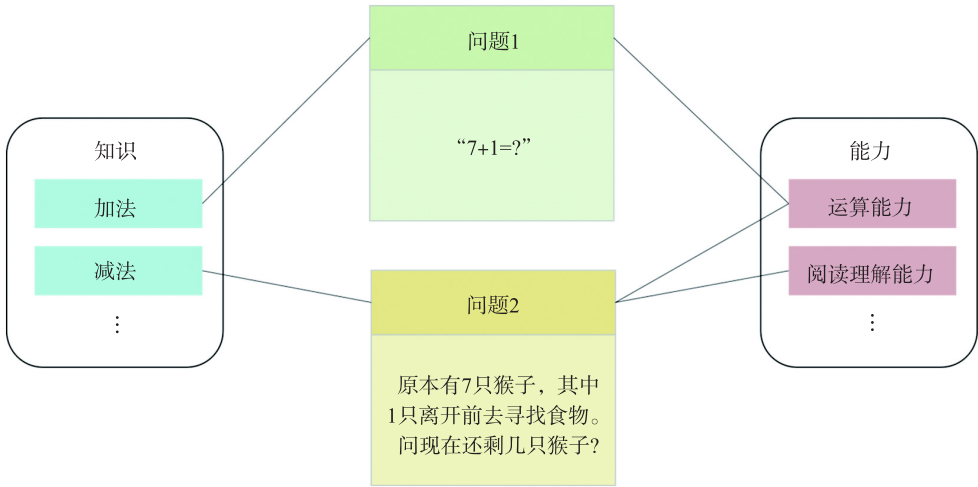


图 1 在实际答题过程中的两个关键因素
Fig. 1 Two key factors in the actual answering process

为了解决这个问题,本文提出了融合知识与能力的深度知识追踪模型. 主要贡献如下:

- (1)提出基于能力的答题预测网络,从能力层面对学生进行建模并预测答题结果.
- (2)提出知识与能力的融合模块,通过将知识和能力两个层面的预测结果进行线性组合,使得模型能够融合两者信息,从而更精准地预测学生的答题结果.
- (3)在 4 个数据集上进行了实验,并与一系列已有的工作进行了对比分析,验证了本文方法的有效性.

1 相关研究

现有的知识追踪方法大致可分为两类:基于传统机器学习的知识追踪和基于深度学习的知识追踪. 本节对这两个方面的主要相关工作进行介绍.

1.1 基于隐马尔可夫模型的知识追踪

1994 年,Corbett 等首次提出了 KT 的概念. KT 可以看作是根据学生过去的学习记录,对学生的知识状态进行动态建模的过程^[1,4]. Corbett 等最早开始 KT 的研究,提出了基于隐马尔科夫模型的 BKT (bayesian knowledge tracing) 模型. BKT 使用二元变量对学生的知识状态进行建模,代表是否掌握某个知识点,并在每次答题结束后使用隐马尔科夫模型动态更新学生的知识. 随后又出现了各种 BKT 的变体. Baker 等^[7]在模型中考虑了学生猜测答案的行为对预测结果的影响;Pardos 等^[8]在建模时引入了题目难度的信息;Käser 等^[9]将动态贝叶斯网络运用于 KT 上,能够关注到不同知识点之间的关联性.

然而,这些基于隐马尔科夫模型的 KT 方法存在一定的缺陷:这些方法片面地将学生对各个知识点的掌握程度分为“掌握”和“未掌握”,不符合现实中循序渐进的学习过程^[10],难以持续性地跟踪学生的知识状态.

1.2 基于深度学习的知识追踪

深度学习以其强大的特征提取能力受到了研究者的广泛关注,研究者提出了许多基于深度学习的知识追踪模型. 2015 年, Piech 等^[3]首次将深度学习运用到 KT 领域,提出 DKT(deep knowledge tracing)模型. 相比于 BKT, DKT 以循环神经网络(recurrent neural network, RNN)为基础,将隐向量表示为学生对各知识点的掌握情况,并能够捕捉到更为复杂的学生知识表征. 这种优势使得基于 RNN 及其变体^[11-12]的模型成为 KT 领域使用最为广泛的模型. 随后出现了基于 DKT 的各种变体^[13-14],相较于 DKT 取得了一定程度的性能提升. 随后出现了基于记忆矩阵的 DKVMN^[4]和基于自注意力的 SAKT^[5]、SAINT^[15]、LANA^[16]等一系列 KT 模型.

然而,这些模型只根据学生对知识点的掌握程度进行预测. 事实上,与具体知识点无关的学生个人能力在答题过程中也发挥着重要的作用. Liu 等^[17]提出的方法虽然考虑了学生的能力因素,但是无法泛化到未采集到答题记录的学生. Zhou 等^[16]提出的方法只将具有相似能力水平的学生划分在同一群体内,无法关注到群体内部各学生之间的细微差异. 本文的方法融合了知识和能力两方面的信息,可以有效适用于没有答题记录的学生,且能够针对每个学生单独进行能力层面的建模,关注到学生之间的差异性.

2 主要方法

本节先介绍知识追踪的定义,然后介绍融合知识与能力的深度知识追踪模型 FKA-DKT 的整体结构,接着分别介绍 FKA-DKT 中的(1)基于知识的答题预测网络(knowledge-based answer prediction network, KAPN)模块,(2)基于能力的答题预测网络(ability-based answer prediction network, AAPN),(3)融合模块,最后介绍分步训练策略.

2.1 问题定义

在现实的学习过程中,学生接触到的各类在线或离线的学习资源都遵循着循序渐进的性质^[18]. 由于很难在这样的学习过程中直接衡量学生的知识状态,因此 KT 模型通常采用一种替代的解决方案(如图 2 所示):根据学生的历史答题记录,预测学生答对下一道题的概率^[1-2]. 学生的知识和能力是预测过程中的两个关键因素,因此本节从知识和能力两个层面介绍知识追踪的定义.

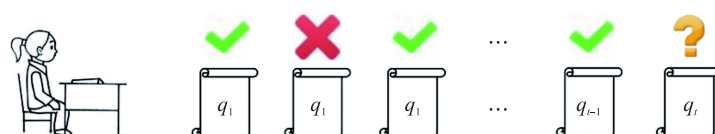


图 2 知识追踪的定义

Fig. 2 Definition of knowledge tracing

2.1.1 知识层面的预测过程

知识层面的预测过程定义如下:给定学生 s 的历史答题记录 $X = \{x_1, x_2, \dots, x_{t-1}\}$, 预测学生答对题目的 q_t 的概率 $P(a_t = 1 | X, q_t)$; $x_i = (q_i, a_i)$ 表示学生在时刻 i 的一次答题记录; q_i 表示时刻 i 的题目知识点编号; $a_i \in \{0, 1\}$ 表示学生的答题结果,其中 0 表示错误,1 表示正确.

2.1.2 能力层面的预测过程

学生的个人能力(例如计算能力、阅读理解能力、记忆能力、推理能力等)在答题过程中起到一定的帮助. 每项能力都可以用一个实数表示. 实向量 \mathbf{A}_i 表示学生在时刻 i 的各项能力水平,称作能力向量. 因此能力层面的知识追踪过程定义如下:给定能力向量 \mathbf{A}_{t-1} , 预测 $P(a_t = 1 | \mathbf{A}_{t-1}, q_t)$; 其中 $\mathbf{A}_{t-1} \in \mathbb{R}^{H_a}$, H_a 表示能力的种类数.

2.2 整体结构

FKA-DKT 模型的整体结构如图 3 所示,由 3 个部分组成,分别是知识模块 KAPN、能力模块 AAPN 和融合模块. KAPN 根据历史答题记录 $\{x_1, x_2, \dots, x_t\}$, 从知识层面预测学生在时刻 2 到 t 答题正确的概率

$\{p_k^2, p_k^3, \dots, p_k^t\}$; AAPN 根据历史答题记录 $\{x_1, x_2, \dots, x_t\}$, 从能力层面预测学生在时刻 2 到 t 答题正确的概率 $\{p_a^2, p_a^3, \dots, p_a^t\}$; 最后, 由融合模块将知识和能力两个层面的预测结果进行线性组合, 计算出最终的预测结果 $\{p^2, p^3, \dots, p^t\}$.

2.3 知识模块: 基于知识的答题预测网络 (KAPN)

KAPN 从知识层面预测学生的答题结果, 其结构如图 4 所示. 首先, 将历史答题记录 $\{x_1, x_2, \dots, x_{t-1}\}$ 转化为独热向量 $\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_{t-1}\}$, 通过式(1)计算出 x_i 的独热向量 \tilde{X}_i 中取值为 1 的位置下标 d_i :

$$d_i = q_i + n \times (1 - a_i), \quad (1)$$

式中, n 为知识点的数量. 然后将独热向量输入至 KAPN 中, 通过嵌入层^[19]转为稠密向量序列 $\{e_1, e_2, \dots, e_{t-1}\}$, 如式(2)所示:

$$e_i = W_{ex} \tilde{X}_i. \quad (2)$$

接着, 使用长短期记忆网络^[11] (long short term memory, LSTM) 根据稠密向量序列 $\{e_1, e_2, \dots, e_{t-1}\}$ 计算学生在不同时刻的知识状态, 如式(3)、(4)所示:

$$h_i = \tanh(W_{he} e_i + W_{hh} h_{i-1} + b_h), \quad (3)$$

$$y_i = \sigma(W_{yh} h_i + b_y), \quad (4)$$

式中, \tanh 和 σ 是激活函数; 下标 i 是当前时刻的时间戳; $W_{ex}, W_{he}, W_{hh}, b_h, W_{yh}, b_y$ 均为可学习的参数.

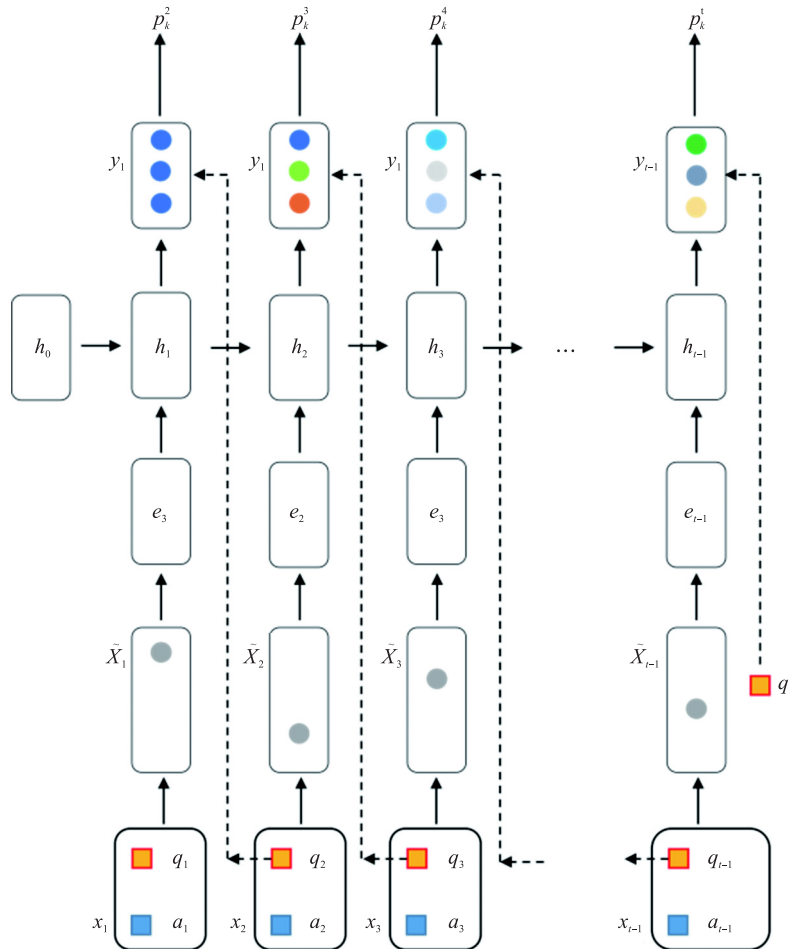


图 4 KAPN 结构图

Fig. 4 Framework of KAPN

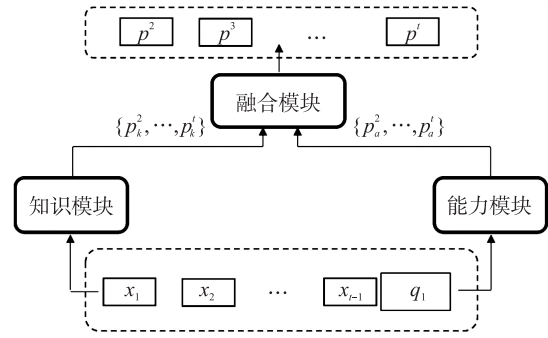


图 3 FKA-DKT 结构图

Fig. 3 Framework of FKA-DKT

$y_i \in [0,1]^n$ 表示学生在时刻 i 对各个知识点的掌握情况. 最后, p_k^i 由式(5)计算得到:

$$p_k^i = y_{i-1, q_i}, \quad (5)$$

式中, y_{i-1, q_i} 为 y_{i-1} 中的第 q_i 个元素.

2.4 能力模块:基于能力的答题预测网络(AAPN)

学生的个人能力在答题过程中也起到重要的作用. 为了追踪学生个人能力的变化情况, 本文提出基于能力的答题预测网络(AAPN), 从能力层面预测学生的答题结果并对学生的个人能力进行建模. 其结构图如图5所示. 下面分别介绍使用AAPN进行能力建模和答题预测的过程.

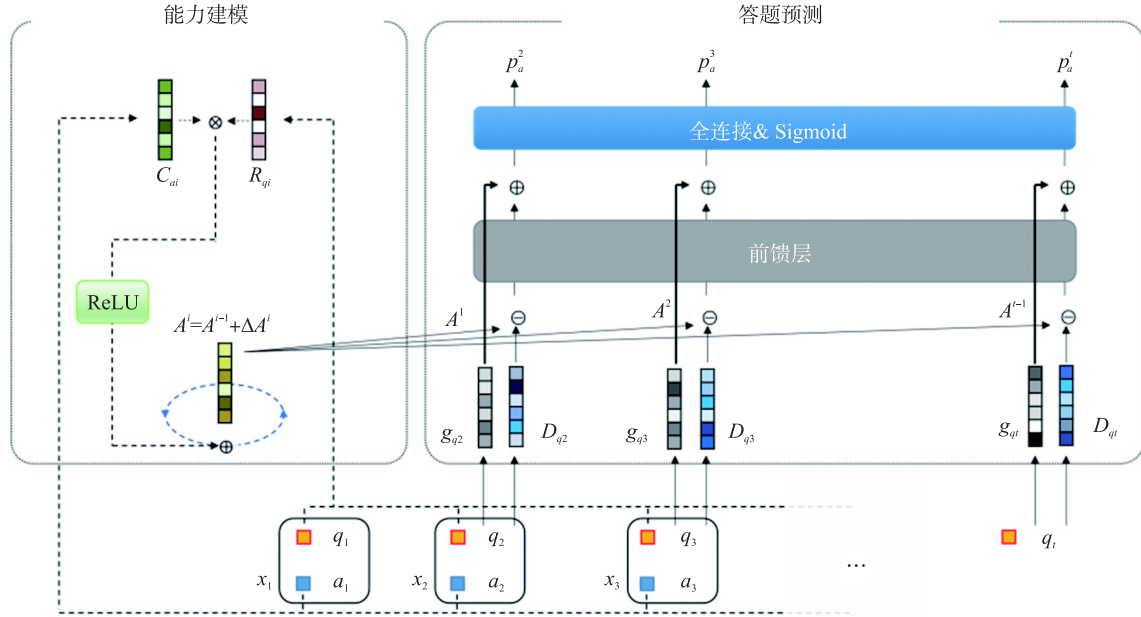


图5 AAPN 结构图

Fig. 5 Framework of AAPN

2.4.1 能力建模

一般来说, 学生在各项能力上(如计算能力、阅读理解能力、记忆能力和推理能力)都具有一定的能力水平, 且这些能力水平会随时间不断变化, 同时这些能力在建模过程中也是不可忽视的因素. 本小节介绍 AAPN 中的能力建模过程, 流程图如图6所示.

将 $A_s^i \in \mathbb{R}^{H_a}$ 定义为学生 s 在时刻 i 回答完问题 q_i 后的能力向量, ΔA_s^i 定义为回答完问题 q_i 后的能力增量, 其中 H_a 是能力的数目. 因此能力向量从时刻 $i-1$ 到 i 的变化过程可以用式(6)来描述:

$$\Delta A_s^i = A_s^i - A_s^{i-1}. \quad (6)$$

特别地, 在学生开始回答问题之前, 其初始能力向量为 $A_s^0 \in \mathbb{R}^{H_a}$.

为了计算学生 s 在任意时刻的能力增量 ΔA_s^i , 本文建立如下两个假设:

(1) 每名学生都拥有一个相等的初始能力向量, 即 $A_s^0 \in \mathbb{R}^{H_a}$ 由全体学生共享.

由假设1可知, 通过 A_s^0 可以对未采集过答题记录的学生进行建模. $A_s^0 \in \mathbb{R}^{H_a}$ 是可学习的参数, 代表了全体学生的平均能力.

(2) 在每道题作答结束后, 学生会根据反馈结果

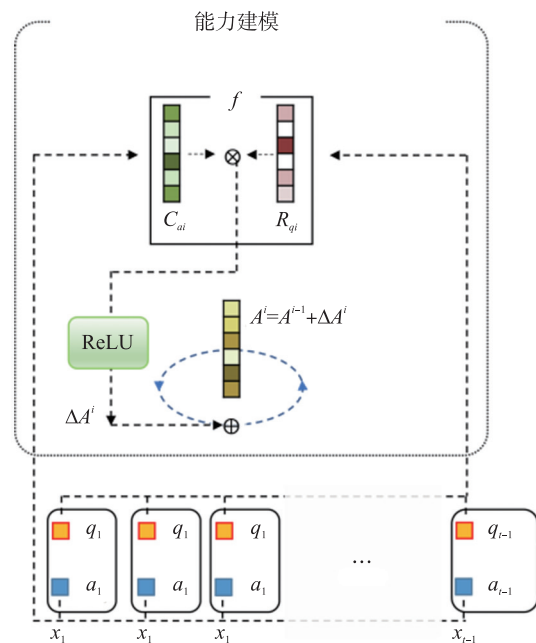


图6 能力建模流程图

Fig. 6 Flow chart of ability modeling

a_i 更新自己在各项能力上的水平. 每项能力的更新幅度均不相同.

由假设 2 可知,能力增量 ΔA_s^i 是以 a_i 和 R_{q_i} 为自变量的二元函数 f , f 称为能力增量函数 (ability increment function, AIF), $R \in \mathbb{R}^{n \times H_a}$ 是可学习的参数, n 是题目的数量;此外,在答题过程中学生的个人能力通常不会呈现下降的趋势,答题总是能够为学生的能力水平带来正反馈^[17],因此能力增量 ΔA_s^i 是一个非负实向量,计算过程如式(7)所示:

$$\Delta A_s^i = \text{relu}(f(a_i, R_{q_i})). \quad (7)$$

针对式(7)中的 f ,本文设计了 f 的两种解析式:

(1) 学生在不同的能力上获取不同的更新值,因此 f 的解析式如式(8)所示:

$$f(a_i, R_{q_i}) = C_{a_i} \circ R_{q_i}, \quad (8)$$

式中, \circ 是向量的位乘, $C \in \mathbb{R}^{2 \times H_a}$ 是可学习的参数,称为能力获取系数矩阵.

(2) 回答正确时的能力增量为 R_{q_i} , 回答错误时的能力增量为 0. 因此 f 的解析式如式(9)所示:

$$f(a_i, R_{q_i}) = a_i * R_{q_i}, \quad (9)$$

式中, $*$ 是向量的数乘.

本文的实验采用公式(8)作为 AIF,并在 3.7 节中探究了两种 AIF 的优劣.

2.4.2 答题预测

在现实的学习过程中,每一道问题都对学生的各项能力有着一定的要求;此外,每道题还包含着抽象的全局能力特征. 学生在能力层面能否答对问题取决于学生自身的能力、题目对学生能力的要求和题目自身的全局能力特征. 本节具体介绍 AAPN 中的答题预测过程.

答题预测的流程图如图 7 所示. 时刻 i 的预测结果取决于学生 s 在时刻 $i-1$ 的能力向量 A_s^{i-1} , 题目对学生的能力要求 D_{q_i} 和全局能力特征 g_{q_i} , 其中 $D \in \mathbb{R}^{n \times H_a}$ 和 $g \in \mathbb{R}^{n \times H_p}$ 都是可学习的参数. 因此 AAPN 的答题预测过程如式(10)所示:

$$p_a^i = \sigma \{ W_{(1, H_p)} [g_{q_i} + \text{ffn}(A_s^i - D_{q_i})] + b_1 \}, \quad (10)$$

式中, H_p 是可调节的超参数. $\text{ffn}(\cdot)$ 是前馈层,对 $A_s^i - D_{q_i}$ 进行抽象化处理,其计算过程如式(11)所示:

$$\text{ffn}(x) = W_{(H_p, m)} [\text{relu}(W_{(m, H_a)} x + b_m)] + b_{H_p}, \quad (11)$$

式中, m 的值设为 2 048^[20].

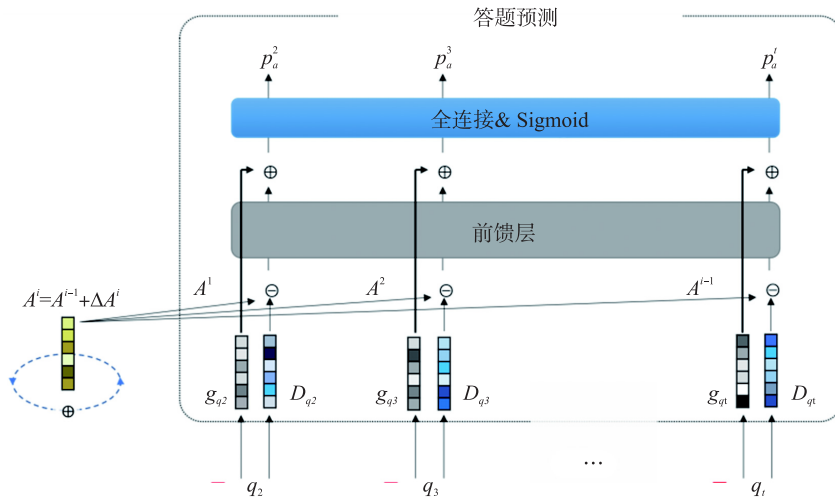


图 7 答题预测流程图

Fig. 7 Flow chart of answering prediction

2.5 融合模块

为了能够融合知识与能力两个层面的信息,使模型更加精准地预测学生的答题结果,本文使用融合模块(如图 8 所示)将 KAPN 和 AAPN 的预测结果进行线性组合,如式(12)所示:

$$p^i = \lambda * p_k^i + (1 - \lambda) * p_a^i, \quad (12)$$

式中,下标 i 表示时间戳; p_k^i 和 p_a^i 分别是 KAPN 和 AAPN 的预测结果, $\lambda \in (0, 1)$ 是平衡系数,用于平衡

KAPN 和 AAPN 的预测结果.

2.6 分步训练策略

平衡系数的存在导致在训练过程中 KAPN 和 AAPN 的梯度绝对值较小,因此直接训练整个模型难以获得较好的效果. 针对这个问题,本文提出分步训练的策略:分别单独训练 KAPN 和 AAPN,在测试阶段加载已训练好的权重,使得模型表现出比直接训练更好的性能. 后续的 3.6 节的实验结果验证了这一策略的有效性.

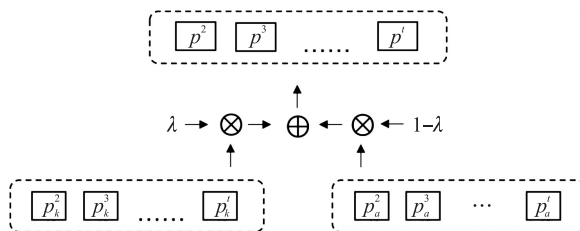


图 8 融合模块结构图

Fig. 8 Framework of fusion module

3 实验结果与分析

本节首先介绍实验所用的数据集、对比方法和评价指标,接着介绍 FKA-DKT 与一系列对比方法的性能对比实验. 最后介绍消融实验、参数选择实验以及可视化分析. 本节中所有的实验数据均取 5 次实验的均值和标准差.

3.1 数据集和对比方法

本文在知识追踪领域 4 个常用的数据集上进行实验,分别是 ASSIST09, ASSIST15, ASSIST17 和 Statics2011. ASSIST 系列数据集由 ASSISTments 在线辅导平台收集,是目前知识追踪领域最常用的基准数据集;Statics2011 收集于某大学静力学课程的辅导系统.

参照现有研究工作,本文对数据集进行预处理. 对于所有数据集,删除其中作答结果不为 1 或 0 的答题记录;对于包含多个知识点的题目,将多个知识点的组合看作一个新知识点;为了避免过参数化的问题^[6],将涉及相同知识点的题目看作同一道题;按照 8:1:1 的比例将数据集划分为训练集、验证集和测试集. 预处理后的数据集统计信息如表 1 所示.

本实验选用 5 个已有的 KT 模型作为对比方法,在 4 个数据集上进行性能对比实验,从而验证本文方法的有效性. 具体的对比方法介绍如下:

(1)DKT^[3]:将独热向量作为输入,使用 LSTM 预测学生的答题结果. 在预测过程中用一个高维向量表示学生的知识状态,向量的维度等于知识点的数目.

(2)DKVMN^[4]:使用两个矩阵存储答题过程中的相关信息,并通过读写操作更新矩阵中的值来模拟学生学习过程中的知识状态变化.

(3)SAKT^[5]:使用自注意力机制预测学生的作答结果,并通过注意力权重提供一定的可解释性.

(4)SAINT^[15]:基于 transformer^[20]模型,将问题和答案的信息分别输入至 encoder 和 decoder 中,并由 decoder 输出预测结果.

(5)LANA^[16]:基于 transformer 模型,在建模过程中考虑了学生的个性化能力,将学生按照能力划分为若干组,并为每一组分配不同的模块权重.

表 1 预处理后的数据集统计信息
Table 1 Statistics of preprocessed datasets

统计信息	数据集			
	ASSIST09	ASSIST15	ASSIST17	Statics2011
题目数	124	100	102	1 224
学生数	4 217	19 840	1 709	333
答题总次数	401 756	683 801	942 816	261 947
平均每个学生的回答次数	95.3	34.5	551.7	786.6
平均每道题的练习次数	3 240.0	6 838.0	9 243.3	214.0

3.2 评价指标

知识追踪任务可以看成是一个二分类问题,即预测学生回答当前时刻题目的正确性(正确或错误). 因此,参照绝大多数现有工作,本文使用 AUC^[21]作为衡量模型预测性能的指标. AUC 定义为 ROC 曲线与下坐标轴转成的面积,是一个值在 0 到 1 之间的实数. AUC 的值越高,表明模型的预测效果越好.

3.3 实验设置

考虑到运行效率问题,在加载数据集时,将 ASSIST15 的最大序列长度设置为 50,其余 3 个数据集的最大长度设置为 100;对于长度小于最大长度的序列,参照现有工作的处理方法^[4],在序列右侧的空白处补零;对于长度大于最大长度的序列,将这样的序列拆分为多个互有重叠的子序列(如图 9 所示),从而保持拆分后学生知识状态的连续性.在拆分过程中,重叠部分的长度为最大长度的三分之一.

模型的超参数设置为:KAPN 的隐层维度为 200;AAPN 中的 H_a, H_p 和融合模块中的平衡系数 λ 均通过超参数搜索确定(在 3.9 节介绍);在模型训练阶段,所选择的优化器为 Adam^[22],学习率设置为 0.001, batch 大小设置为 128. 实验中使用的损失函数为二元交叉熵(binary cross entropy, BCE)损失,对于单个学生而言,损失计算过程如式(13)所示:

$$\text{BCE}(\{a_2, a_3, \cdots, a_t\}, \{y_2, y_3, \cdots, y_t\}) = - \sum_{i=2}^t (a_i(y_i) + (1-a_i)(1-y_i)), \tag{13}$$

式中,下标 i 为时间戳, a_i 为该学生在时刻 i 的真实作答结果, y_i 为模型的预测值,式(13)中所有的对数函数 $\log(\cdot)$ 均以 2 为底;对于整个数据集而言,将所有学生的 BCE 损失值相加即为最终的损失函数.

3.4 对比实验

表 2 展示了各个模型在 4 个数据集上的 AUC 值. 相较于次优的对比方法, FKA-DKT 在 4 个数据集上分别以 1.2%, 0.8%, 1.0% 和 1.3% 的相对优势占据最优地位. 此外,相较于没有考虑学生能力的 DKT, DKVMN, SAKT 和 SAINT, FKA-DKT 综合考虑了学生的能力变化过程与自适应性,能够将能力因素融入进知识追踪中,取得更为准确的预测结果;相较于将学生分组并考虑群体能力的 LANA, FKA-DKT 则更为细致地关注到每个学生之间的能力差异性,能够为每个学生单独进行能力层面的建模,这表明在知识追踪中,不同学生之间的能力差异性值得关注,以学生为中心的能力建模方法优于以群体为中心的能力建模方法.

表 2 FKA-DKT 与对比模型的对比实验
Table 2 Performance comparison between FKA-DKT and baselines

数据集	AUC/%					
	DKT	DKVMN	SAKT	SAINT	LANA	FKA-DKT
ASSIST09	79.91±0.21	77.02±0.10	75.85±0.10	76.90±0.15	79.95±0.15	80.93±0.18
ASSIST15	72.23±0.06	70.40±0.04	70.99±0.08	69.74±0.46	72.06±0.09	72.84±0.06
ASSIST17	69.92±0.14	66.31±0.15	64.26±0.06	67.52±0.18	70.35±0.27	71.08±0.11
Statics2011	84.26±0.26	80.59±0.24	77.87±0.09	78.67±0.28	83.61±0.26	85.36±0.20

3.5 消融实验

为了探究知识和能力两个因素在知识追踪中的影响,本实验分别使用知识模块 KAPN,能力模块 AAPN 和 FKA-DKT 在 4 个数据集上进行性能对比,实验结果如图 10 所示. 图中结果表明在 4 个数据集上, KAPN 在预测学生答题表现文献普遍优于 AAPN,然而效果最佳的模型还是能够将知识与能力融合的 FKA-DKT. 这表明在现实的学习过程中,知识与能力都发挥着重要作用,因此考虑能力因素是必要的;此外,仅考虑能力因素预测学生表现的效果不如仅考虑知识的情况,这说明知识因素在学习过程中相对更重要,这与现实经验一致.

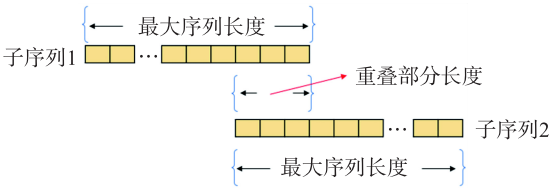


图 9 序列拆分示意图
Fig. 9 Diagram of sequence splitting

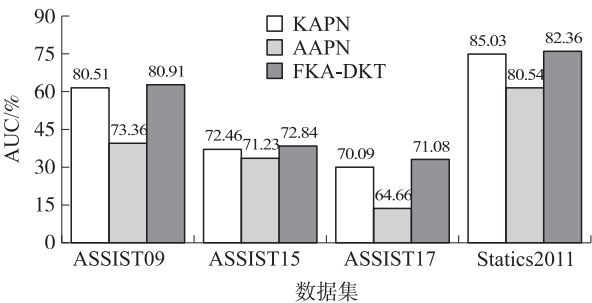


图 10 KAPN, AAPN, FKA-DKT 在 4 个数据集上的性能对比
Fig. 10 Performance comparison between KAPN, AAPN, and FKA-DKT on 4 datasets

3.6 分步训练策略

本节实验探究了分步训练策略对模型性能的影响. 直接训练和分步训练的模型性能对比如表 3 所示. 分析表中数据可知:相较于直接训练整个模型,使用分步训练策略训练模型普遍有助于模型预测效果的改善;此外,在 ASSIST 系列数据集上使用分步训练策略还能使模型以更小的标准差稳定地占据性能优势.

3.7 AIF 的解析式

本节实验对比了分别采用式(8)和式(9)作为 AIF 的模型性能,实验数据如表 4 所示. 由实验数据可知:式(8)的 AIF 能够更好地对学生的个人能力进行建模. 这意味着无论作答结果是正确还是错误,学生都能在各项能力上取得一定程度的正反馈,而使用式(8)的 AIF 能够较好地模拟这一过程;相比之下,式(9)的 AIF 对模型造成了极大的负面影响,这是因为式(9)的 AIF 认为学生回答错误时不会取得任何能力正反馈,这样的建模过程过于片面,无法反映真实的学习情况.

表 3 分步训练和直接训练的性能对比

Table 3 Performance comparison between stepwise training and directly training

数据集	AUC/%	
	直接训练	分步训练
ASSIST09	80.63±0.27	80.93±0.18
ASSIST15	72.14±0.13	72.84±0.06
ASSIST17	70.76±0.16	71.08±0.11
Statics2011	85.11±0.12	85.36±0.20

表 4 分别采用两种 AIF 的模型性能

Table 4 Performance comparison between 2 AIFs

数据集	AUC(%)	
	(8) $C_{a_i} \circ R_{q_i}$	(9) $a_i * R_{q_i}$
ASSIST09	80.93±0.18	49.76±1.69
ASSIST15	72.84±0.06	51.77±1.76
ASSIST17	71.08±0.11	49.56±0.94
Statics2011	85.36±0.20	52.67±1.48

3.8 知识与能力的可视化分析

为了进一步探究 FKA-DKT 的决策过程,本文以 ASSIST09 和 ASSIST17 为例,对知识和能力层面的预测结果进行可视化分析. 图 11 展示了两个学生在与特定技能相关的问题上的学习序列,底部横坐标表示时刻,纵坐标表示知识/能力层面的预测概率,顶部横坐标表示对应时刻的实际作答结果. 从图中可以看出,图中的知识掌握度曲线与实际作答结果基本吻合;此外,能力匹配度的变化也有效地验证了本文的结论. 以图 11(a)中的学生为例,该学生在回答涉及“参数估计”的问题时,与“参数估计”相关的各项能力(如计算能力、推理能力等)在不断地提升;在图 11(b)中,即使该学生连续多次答错同一道题,但其能力匹配度趋于稳定,这表明在实际的学习过程中,学生的能力水平通常不会下降,这与平时的生活经验一致. 这两个例子展示了 FKA-DKT 在模型可解释性方面的优势,并为学生学习过程中的反馈提供不同的视角.

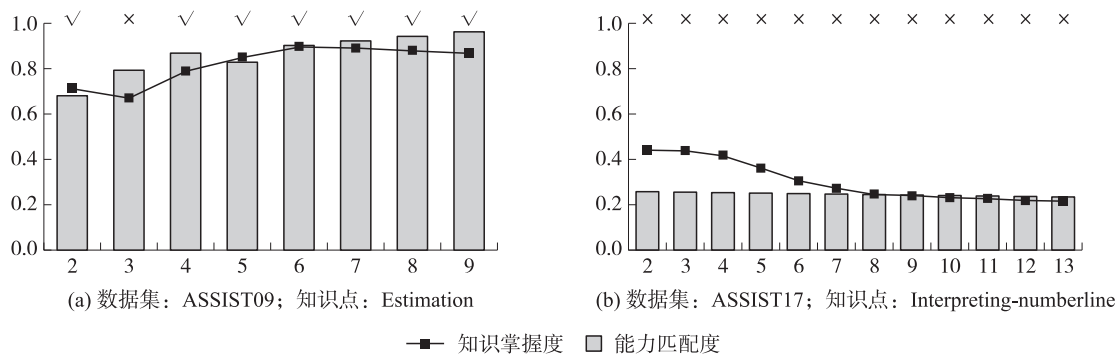


图 11 知识与能力层面预测结果的可视化分析

Fig. 11 Visualization analysis of prediction on knowledge and ability aspects

3.9 超参数选择实验

为了充分探究在不同情况下,模型的预测性能受超参数的影响状况,本实验选取了部分超参数取值组合并记录了在这些超参数取值下模型的 AUC 值,如表 5 所示. 从表中可知:

(1) H_a 的取值与数据集中的题目数量成非严格正相关. ASSIST09, ASSIST15 的题目数量较少,因此可以选取 $H_a = 100$; Statics2011 的题目数量较多,因此可以选取 $H_a = 150$;而在 ASSIST17 上,可能由于偏少的学生数目和偏多的答题次数导致每个学生的平均序列长度远远大于 ASSIST09 和 ASSIST15,从而导致

ASSIST17 中过多地包含学生自身的各种意外行为(失误、猜测、遗忘、思维定势等),影响到 H_a 的最佳取值.

(2) H_p 的取值与数据集的规模有关,由于 Statics2011 的答题总次数明显小于其它 3 个数据集,因此在 Statics2011 上选取一个较小的 H_p 能够达到模型最佳性能.

(3)在 ASSIST15 和 ASSIST17 上, λ 的最优取值为 0.7;在 ASSIST09 和 Statics2011 上, λ 的最优取值为 0.8. 这意味着在学习过程中学生的个人能力通常占据着 20%~30%的比例,是一个不容忽视的因素.

表 5 FKA-DKT 在 4 个数据集上的超参数选择实验结果,mean 和 std 是 AUC(%)的均值和标准差
Table 5 FKA-DKT hyperparameter selection experimental results on 4 datasets,mean and std are the mean value and standard deviation of AUC(%)

ASSIST09							ASSIST15						
H_a	75	100	100	100	100	100	75	100	100	100	100	100	100
H_p	128	128	256	256	256	512	128	128	256	256	256	512	512
λ	0.7	0.7	0.7	0.6	0.8	0.7	0.7	0.7	0.7	0.6	0.8	0.7	0.7
mean	80.82	80.81	80.91	80.51	80.93	80.73	72.75	72.84	72.84	72.80	72.77	72.76	72.76
std	0.16	0.13	0.08	0.12	0.18	0.18	0.07	0.10	0.06	0.08	0.11	0.06	0.06

ASSIST17							Statics2011						
H_a	125	150	150	150	150	150	125	150	150	150	150	150	150
H_p	128	128	256	512	256	256	256	256	128	64	64	64	64
λ	0.7	0.7	0.7	0.7	0.8	0.6	0.7	0.7	0.7	0.7	0.8	0.6	0.6
mean	71.00	70.95	71.08	70.92	71.03	70.97	85.31	85.31	85.21	85.35	85.36	85.17	85.17
std	0.13	0.22	0.11	0.06	0.16	0.21	0.16	0.11	0.16	0.17	0.20	0.12	0.12

4 结论

针对现有深度知识追踪模型在预测过程中普遍缺乏考虑学生个人能力因素的问题,本文提出了融合知识与能力的深度知识追踪模型,从知识和能力两个层面预测学生的答题结果,最后将两个层面的预测结合进行线性组合,使模型能够融合知识和能力两个层面的信息预测学生的答题情况. 在 4 个公开数据集上的实验结果验证了本文方法的有效性.

当前的大部分 KT 工作忽略了涉及相同知识点的题目之间的差异性和很多隐藏在历史答题记录中的侧面信息^[23](例如题目类别、模板、学生答题用时、是否求助等),而这些信息尚未受到研究者的广泛关注. 如果能充分利用这些信息,模型的性能可以得到进一步的提升,而这也是 KT 领域未来的研究方向.

[参考文献]

[1] CORBETT A T,ANDERSON J R. Knowledge tracing:modeling the acquisition of procedural knowledge[J]. User modeling and user-adapted interaction,1994,4(4):253-278.

[2] 刘铁园,陈威,常亮,等. 基于深度学习的知识追踪研究进展[J]. 计算机研究与发展,2022,59(01):81-104.

[3] PIECH C,BASSEN J,HUANG J,et al. Deep knowledge tracing[J]. Advances in neural information processing systems, 2015,28:505-513.

[4] ZHANG J,SHI X,KING I,et al. Dynamic key-value memory networks for knowledge tracing[C/OL]//Proceedings of the 26th International Conference on World Wide Web. Republic and Canton of Geneva,CHE;International World Wide Web Conferences Steering Committee,2017:765-774[2022-02-20]. <https://doi.org/10.1145/3038912.3052580>.

[5] PANDEY S,KARYPIS G. A self-attentive model for knowledge tracing[J/OL]. ArXiv:1907.06837[cs,Stat],2019[2022-02-21]. <http://arxiv.org/abs/1907.06837>.

[6] GHOSH A,HEFFERNAN N,LAN A S. Context-aware attentive knowledge tracing[C/OL]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York,NY,USA;Association for Computing Machinery,2020:2330-2339[2022-02-21]. <https://doi.org/10.1145/3394486.3403282>.

[7] BAKER R S J D,CORBETT A T,ALEVEN V. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing[C]//Intelligent Tutoring Systems. Berlin,Heidelberg:Springer,2008:406-415.

- [8] PARDOS Z A, HEFFERNAN N T. KT-idem: introducing item difficulty to the knowledge tracing model[C]//User Modeling, Adaption and Personalization. Berlin, Heidelberg: Springer, 2011: 243–254.
- [9] KÄSER T, KLINGLER S, SCHWING A G, et al. Dynamic bayesian networks for student modeling[J]. IEEE transactions on learning technologies, 2017, 10(4): 450–462.
- [10] 李晓光, 魏思齐, 张昕, 等. LFKT: 学习与遗忘融合的深度知识追踪模型[J]. 软件学报, 2021, 32(03): 818–830.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735–1780.
- [12] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J/OL]. ArXiv: 1412.3555[Cs], 2014[2022-01-28]. <http://arxiv.org/abs/1412.3555>.
- [13] MINN S, YU Y, DESMARAI S M C, et al. Deep knowledge tracing and dynamic student classification for knowledge tracing[C]//2018 IEEE International Conference on Data Mining(ICDM). IEEE, 2018: 1182–1187.
- [14] YEUNG C K, YEUNG D Y. Addressing two problems in deep knowledge tracing via prediction-consistent regularization[Z/OL]. London, United Kingdom: Association for Computing Machinery, 2018: 1–10 (2018). <https://doi.org/10.1145/3231644.3231647>.
- [15] CHOI Y, LEE Y, CHO J, et al. Towards an appropriate query, key, and value computation for knowledge tracing[C/OL]//Proceedings of the Seventh ACM Conference on Learning@ Scale. New York, NY, USA: Association for Computing Machinery, 2020: 341–344[2022-06-09]. <https://doi.org/10.1145/3386527.3405945>.
- [16] ZHOU Y, LI X, CAO Y, et al. LANA: towards personalized deep knowledge tracing through distinguishable interactive sequences; arXiv: 2105.06266[Z/OL]. arXiv, 2021 (2021-04-20) [2023-03-20]. <http://arxiv.org/abs/2105.06266>.
- [17] LIU S, YU J, LI Q, et al. Ability boosted knowledge tracing[J]. Information sciences, 2022, 596: 567–587.
- [18] GABRIELLA C, GRILLI L, LIMONE P, et al. Deep learning for knowledge tracing in learning analytics: an overview[C]//CEUR Workshop Proceedings. Foggia, Italy: CEUR-WS, 2021: 1–10.
- [19] WANG W, MA H, ZHAO Y, et al. Tracking knowledge proficiency of students with calibrated q-matrix[J]. Expert systems with applications, 2022, 192: 116454.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J/OL]. ArXiv: 1706.03762[Cs], 2017[2021-11-26]. <http://arxiv.org/abs/1706.03762>.
- [21] LING C X, HUANG J, ZHANG H. AUC: a statistically consistent and more discriminating measure than accuracy[C]//International Joint Conference on Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 2003: 519–524.
- [22] KINGMA D P, BA J. Adam: a method for stochastic optimization; arXiv: 1412.6980[Z/OL]. arXiv, 2017 (2017-01-29) [2022-06-24]. <http://arxiv.org/abs/1412.6980>.
- [23] LIU Q, SHEN S, HUANG Z, et al. A survey of knowledge tracing[J/OL]. ArXiv: 2105.15106[Cs], 2021[2022-02-16]. <http://arxiv.org/abs/2105.15106>.

[责任编辑: 陆炳新]