

学习者知识追踪研究进展综述



张 暖¹ 江 波²

1 浙江工业大学教育科学与技术学院 杭州 310023

2 华东师范大学教育信息与技术学系(上海数字化教育装备工程技术研究中心) 上海 200062

(353211550@qq.com)

摘 要 学习者建模是自适应学习系统的支撑技术之一,其中以知识追踪为代表的学习者知识状态建模研究最为广泛。3种代表性的知识追踪技术分别为基于隐马尔可夫模型的贝叶斯知识追踪、基于逻辑回归模型的可加性因素模型、基于循环神经网络的深度知识追踪。通过综述发现,贝叶斯知识追踪模型适用于含单一知识点的学习任务的知识追踪,可加性因素模型和深度知识追踪模型适用于含多知识点的学习任务的知识追踪,但深度知识追踪模型的教学可解释性不佳。在综述现有研究的基础上,受到知识空间理论的启发,将知识点之间的先决关系融入到知识追踪模型是未来的一个重要研究方向,并初步提出了一种融合知识点先决关系的可加因素模型。

关键词: 自适应学习系统;知识追踪;贝叶斯知识追踪;可加性因素模型;深度知识追踪;知识空间理论

中图法分类号 TP391.6

Review Progress of Learner Knowledge Tracing

ZHANG Nuan¹ and JIANG Bo²

1 College of Educational Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

2 Department of Educational Information and Technology(Shanghai Engineering Research Center of Digital Education Equipment), East China Normal University, Shanghai 200062, China

Abstract Learner modeling is one of the supporting techniques of adaptive learning systems, among which learner knowledge state modeling represented by knowledge tracing is the most widely studied. Three representative knowledge tracing techniques are Bayesian Knowledge Tracing(BKT) based on hidden Markov model, Additive Factor Model(AFM) based on logistic regression model and Deep Knowledge Tracing(DKT) based on recurrent neural network. It is found that the BKT is suitable for knowledge tracing in learning tasks that only contain single knowledge skill, AFM and DKT can be used for tracing students' knowledge state in learning tasks that have more than one knowledge skills. However, the DKT is hard to be interpreted from the perspective of pedagogy. Based on reviewing the existing research and inspired by the knowledge space theory, this paper argues that the integration of the prerequisite relationship among knowledge skills and the knowledge tracing is a promising research direction. Finally, this paper proposes a prototype of additive factor model integrating knowledge prerequisite relationship.

Keywords Adaptive learning system, Knowledge tracing, Bayesian knowledge tracing, Additive factor model, Deep knowledge tracing, Knowledge space theory

1 前言

学习者建模(learner modeling)是指利用学习者与数字化学习系统的学习交互数据,对学习者的知识状态、学习情绪和学习行为等内隐学习特征进行刻画,是构建自适应学习系统的核心^[1]。学习者知识状态建模,又称知识追踪(knowledge

tracing),是研究最早和最受关注的一类学习者建模技术。知识追踪根据学习者在多个历史学习任务上的正误序列来预测对学习任务所含知识点的掌握概率,以实现对学生知识掌握状态的动态追踪。自适应学习系统通过对学习者的精准知识追踪,来实现自动构建针对性的教学支架,生成适应性的学习路径和推送个性化的学习资源,实现“因

到稿日期:2020-06-05 返修日期:2020-10-09 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61977058);上海市“科技创新行动计划”人工智能科技支撑专项(20511101600);中央高校基本科研业务费项目华东师范大学引进人才启动费项目(41300-20101-222696)

This work was supported by the National Natural Science Foundation of China(61977058),“Science and Technology Innovation Action Plan” Artificial Intelligence Technology Support Special Project of Shanghai(20511101600) and Fundamental Research Funds for the Central Universities of East China Normal University(41300-20101-222696).

通信作者:江波(bjiang@deit.ecnu.edu.cn)

材施教”,提高学习效果与效率^[2]。

知识追踪的一般思想为,给定学习者在含有某特定知识点的一组学习任务上的正误序列 x_1, x_2, \dots, x_t (t 为回答问题的次数),预测学习者下一次答题正误 x_{t+1} 的概率,并以此估计学习者的知识点掌握状态。从机器学习的角度来看,知识追踪是根据含噪声的观测数据(正误序列)估计隐含变量(知识掌握),是一类典型的预测问题。国内外学者对知识追踪进行了充分而又深入的研究,取得了一系列代表性的研究成果。根据所用机器学习方法的不同,现有研究可大致分为 3 类:基于隐马尔可夫模型的贝叶斯知识追踪(Bayesian Knowledge Tracing, BKT)^[3]、基于逻辑回归模型的可加性因素模型(Additive Factor Model, AFM)^[4]以及基于循环神经网络的深度知识追踪(Deep Knowledge Tracing, DKT)^[5]。BKT 将学习者的学习过程视作隐马尔可夫过程,将历史成绩作为观测变量,将知识点掌握与否作为状态变量,并考虑猜测和失误概率,利用贝叶斯概率公式计算在当前观测变量下学习者的知识掌握概率;AFM 将学习者的能力、知识点难度、学习率等作为参数建立逻辑回归模型;DKT 利用学习者在历史任务上的结果序列,构建长短期记忆(Long Short-Term Memory, LSTM)神经网络模型。

在已有的相关综述研究中,Xu 等^[6]在介绍学习者模型时概述了 BKT 及相关变体以及 DKT 模型。Desmarais 等^[7]回顾了学习者模型,以及在学习者技能建模和评估方面的最新进展,并讨论了其他关键结构建模方面的相关进展,如学习者动机、情感和注意力状态。Pelánek^[8]从建立学习者模型的目的、模型的组成、建模的技术、模型的拟合和评估 4 个方面对 BKT 和 AFM 进行了概述。Li 等^[9]从数据、知识点和学习者 3 个方面总结了 BKT 在教育领域的应用研究现状。Liu 等^[10]总结了知识追踪的相关理论,对现有的方法进行了分类并对技术做了比较。需要指出的是,Liu 等对知识追踪的综述涉及的范围更广,包括了相关教育学理论、数据挖掘技术等,而本文只考虑了 3 类主流的建模方法:BKT, AFM 和 DKT。在此限制下,本文分别从模型原理及模型的改进两方面对相关研究进行更细致的讨论。

现有研究很少将知识追踪与领域模型(Domain Model)相结合进行知识状态预测。领域模型是指知识点的划分、知识点与问题之间的对应关系以及知识点之间的对应关系等模型^[8]。现有的很多知识追踪研究只预测学习者在单个知识点上的掌握状态,在如何利用多个知识点之间的先决关系进行知识状态预测方面,研究尚处于起步阶段。受到知识空间理论(Knowledge Space Theory, KST)^[11]的启发,本文认为将知识点之间的先决关系融入模型中有利于知识状态的预测,是未来知识追踪的研究重点,并提出了初步的研究设想。

2 贝叶斯知识追踪(BKT)

2.1 模型原理

1995 年,Corbett 等将 BKT 模型引入智能教辅系统,用来描述学习者在知识习得过程中知识状态的变化^[12]。BKT 以布鲁姆掌握学习理论^[13]和最近发展区理论^[14]为理论前提,认为领域知识可以分解,学习者在掌握高级知识之前需要

掌握必要的前序知识。BKT 的本质是关于时间序列的隐马尔可夫概率模型,隐马尔可夫链生成不可观测的状态随机序列,由随机状态产生观测随机序列^[15]。在 BKT 中,知识节点的状态为潜变量(不可观测的状态随机序列),成绩节点为可观测变量(观测随机序列)。

BKT 基于以下 4 个假设:学习过程是知识从未掌握状态到掌握状态的离散过渡;知识的状态为掌握与未掌握的二元潜变量;成绩为正确与错误的二元观测变量;学习者不存在遗忘现象。同时,BKT 只对特定的单个知识点进行建模,每个模型相对于知识点独立,不涉及多知识点问题。

如图 1 所示,在经典 BKT 模型中,学习者在每个知识点上的掌握状态主要受到 4 个参数的影响: $P(L)$, $P(T)$, $P(G)$ 和 $P(S)$ 。 $P(L)$ 为先验概率,表示学习者在学习前掌握知识点的初始概率; $P(T)$ 为学习概率,表示通过学习,学习者掌握知识点的转换概率; $P(G)$ 为猜测概率,表示学习者在没有掌握知识点的情况下猜对的概率; $P(S)$ 为失误概率,表示学习者在掌握知识点的情况下答错的概率。系统会根据学习者每次给出的回答持续更新其掌握知识点的概率,一般当 $P(L)$ 达到 0.95 即认为学习者掌握了知识点。

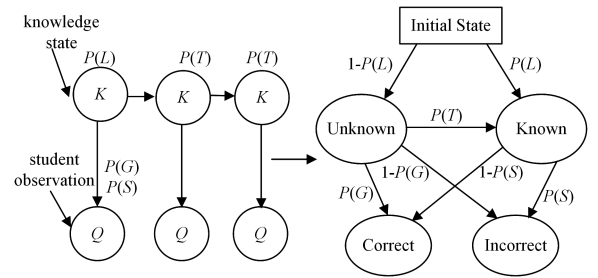


图 1 知识追踪模型图

Fig. 1 Diagram of knowledge tracing model

学习者答对题目的情况有两种:1)掌握了知识点并且没有失误;2)没有掌握知识点但在做题过程中猜对了。因此,学习者答对题目的概率为两者之和,即:

$$P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * P(G) \quad (1)$$

学习者答错题目的情况也有两种:1)掌握了知识点但在做题过程中产生失误;2)没有掌握知识点且在做题过程中猜错了。因此,学习者答错题目的概率为两者之和,即:

$$P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G)) \quad (2)$$

根据式(1)和式(2)可以分别估计学习者在答对和答错两种情况下的知识点掌握概率,分别如式(3)、式(4)所示:

$$P(L_{n-1} | correct) = \frac{P(L_{n-1}) * (1 - P(S))}{P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * P(G)} \quad (3)$$

$$P(L_{n-1} | incorrect) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G))} \quad (4)$$

与测评不同,学习者在学习过程中的知识掌握状态是动态的,学习者可以通过学习教学材料来习得知识点,实现知识掌握状态的转换。BKT 使用学习概率 $P(T)$ 表示学习者的知识掌握状态的转换概率,从而得到式(5)所示的知识状态估计:

$$P(L_n | action_n) = P(L_{n-1} | action_n) + (1 - P(L_{n-1} | action_n)) * P(T) \quad (5)$$

2.2 模型改进

经典 BKT 以贝叶斯概率公式为建模基础,知识追踪过程及结果具有统计学可解释性,但是对于模型的研究与应用而言,尚且存在一定的局限性。因此,研究者做了一系列广泛而深入的研究扩展。下面从模型的假设改进、应用研究和其他研究 3 个方面综述相关的 BKT 研究。

2.2.1 模型的假设改进

改进知识状态假设的研究。经典 BKT 中,知识状态是二元的,即掌握或没有掌握,但是二元状态不能完全反映学习的复杂性。Zhang 等^[16]在原有的二元状态的基础上加入“过渡学习状态”以表示学习者可能掌握知识点的情况,由此将知识状态分为未掌握、可能掌握和掌握 3 种。Agarwal 等^[17]为使学习水平估计的更新更平滑,用近因权重代替学习率,将二元状态进一步细化为 21 个状态。

改进学习结果假设的研究。学习者答题的结果有正确与错误两种情况,因此模型将输入的学习者成绩序列编码为 0 和 1,但是这两种结果无法完全反映学习者的做题过程。Wang 等^[18]在 2010 年提出用连续赋分的方式来代替二元结果,主要依据学习者使用智能提示和练习的次数给学习结果打 0~1 之间的分数,该模型提高了对学习者的预测。但 Wang 等进一步发现,该模型也使得学生在解决问题时必须做更多的练习或要求系统准备更多的提示^[19]。

增加关于时间因素的研究。学习者在在线平台上的学习通常跨越了一段时间,因此,模型需要纳入学习者由于时间跨越产生的遗忘特性^[20]。Qiu 等^[21]对学习者的做题时间间隔较长(一天及以上)的情况进行了研究,确定学习者存在遗忘的现象。修正后的模型加入了遗忘特性,在拟合数据集上表现更优。Khajah 等^[22]在比较 DKT 与 BKT 模型的差异时,为 BKT 加入了时间间隔较短的遗忘概率参数,加入遗忘参数的模型表现出比常规 BKT 模型更好的预测效果。Lyu 等^[23]在模型中加入学生做题的持续时间,以研究每一次做题时间对预测准确率的影响。Zhu 等^[24]通过数据切片算法在成绩数据中检测学习者认知水平变化较大的认知拐点,并提取时间差异信息,以权重的形式将差异信息整合到细化的 BKT 模型中,从而准确量化知识状态。

模型的个性化研究。学习者是具有独立意识的单个的人,BKT 参数的优化可以帮助学习者避免增加额外的练习^[25]。Lee 等^[26]通过实验比较了基于学习者个体的模型和基于学习者群体的模型,发现学习者得到练习机会的次数存在差异。这表明对学习者的参数进行个性化赋值能使模型得到一定的改善,对使用知识追踪模型的智能辅导系统中的教学决策有重要影响。以下研究尝试对每一个学习者参数进行个性化。首先,失误和猜测概率应该随做题的过程不断变化,而不是在所有情况下固定不变。Baker 等^[27]利用机器学习对学生猜测和失误的概率进行动态估计。其次,针对先验概率参数 $P(L)$,Pardos 等^[28]提出学习者优先(Prior Per Student, PPS)模型。该模型为每一个学生赋予不同的先验概率参数,

实现了基础能力的个性化。Nedungadi 等^[29]在 Pardos 的基础上,为每个学习者和知识点提供了单独的先验概率,并尝试对学习者的动态聚类。再次,针对学习参数 $P(T)$,Yudelson 等^[30]提出了一种个性化学习者学习速率的知识追踪模型,通过与学习者优先模型进行比较,发现学习速率个性化的模型比先验概率个性化的模型的预测准确率更高。

在模型中增加多源信息的研究。学习内容本身的特质、学习者的生理特征及认知特点同样会影响知识追踪的精度。针对学习内容的特点,Pardos 等^[31]通过给每一个问题建立差异化的猜测参数和失误参数,引入项目难度,建立了项目难度影响模型。针对学习者的特点,Xu 等^[32]在知识追踪中加入 EEG 脑电波信息以提高知识状态估计,发现在不同脑电状态下学习者的参数具有显著差异。Spaulding 等^[33]在 HMM 中添加额外的可观察节点用于估计学习者的情绪状态,包括困惑、无聊、投入等学习情绪。Lin 等^[34]将教学干预措施加入到 BKT 中,与传统 BKT 相比,该模型在个性化干预方面具有较好的潜力。鉴于模型中可增加信息的多样性,González 等^[35]尝试提出了一类以逻辑回归形式表示每一种参数的通用模型,将一般的特征都整合到知识追踪中。

2.2.2 模型的应用研究

BKT 的应用主要集中于智能教辅系统(Intelligent Tutoring System, ITS)和 MOOC 平台。Beck 等^[36]研究智能教学系统中提示的作用,认为提示有助于学习者长期地学习。此外,MOOC 平台包含大量课程信息,且允许学习者多次提交答案。针对 MOOC 平台的这一数据特点,Pardos 等^[37]先根据提交次数训练 $P(G)$ 和 $P(S)$,并为额外的学习资源设置可观测资源节点 R ,在此限制下将学习参数 $P(T)$ 转换为 $P(T|R)$ 。在 MOOC 中,同一知识点可以对应多个问题,而问题具有不同的难度。Wang 等^[38]利用在时间上最相近的答题序列,即前一次的答题情况分别训练 $P(G)$ 和 $P(S)$,并为每个知识点对应的不同难度的问题设置 Aspect 参数。Pardos 等和 Wang 等的研究更好地利用了不同方面的数据,使得模型更加贴合学习者真实的学习环境,改善了学习系统。Wang 等^[39]针对 Coursera 的课程知识成分之间丰富的结构和相关性,从平台内容组织特点出发对知识状态进行建模,以提高知识状态的建模精度。

2.2.3 其他研究

其他研究包括模型的比较研究和机器学习技术上的改进研究。模型的比较包括集成方法与单个方法的比较^[40]、多类型方法的比较^[41]、BKT 与 DKT 的比较^[22]、BKT 与 IRT 的比较^[42]与结合^[43]以及 BKT 与动态贝叶斯网络的比较^[44]。技术上的改进有利用过采样技术改善类不平衡问题^[45]、建模学习者参数空间的研究^[46]、参数化掌握阈值的研究^[47]、误差度量的比较研究^[48]、简化 BKT 参数估计方法^[49]、研究样本大小的变化对 BKT 模型中参数值的影响程度,以及这些误差对学生的知识掌握程度预测的影响^[50]等。

尽管 BKT 及其扩展在知识追踪领域中被广泛研究与使用,但其一开始对于学习的假设就具有先天的局限性。首先,

知识点的掌握程度是二元的,不允许模糊性的存在,之前的研究并没有很好地解决这个问题。其次,BKT中学习者的回答也是二元的,使得BKT的应用范围受限。最后,BKT只允许单个知识点的建模,需要专家对相关知识进行标记,这需要耗费大量的时间和精力。研究人员在心理测量领域中发现了一类可以解决BKT相关问题的模型,即AFM。

3 可加性因素模型 (AFM)

3.1 模型原理

学习者能力的可测性是心理测量领域的基本观点,项目反应理论(Item Response Theory,IRT)就是建立在这一观点之上。IRT模型在教育评估与测量中发挥着重要的作用。它主要用于分析考试成绩,将潜在的学习者心理特征(能力)通过一系列的考试题目反映出来。IRT认为,主客观因素均可影响学习者的作答成绩。主观因素即学习者自身的能力,客观因素包括题目的难度、区分度、猜对概率和失误概率等。IRT模型基于Logistic函数,通过加入学习者能力、题目难度、题目区分度等参数来预测学习者的成绩。其中一个较全面的IRT模型为4PLM-IRT^[42],如式(6)所示:

$$P(Y_{ij}=1|\theta_i, a_j, b_j, c_j, d_j) = c_j + (d_j - c_j) \frac{e^z}{1 + e^z} \quad (6)$$

$$z = a_j(\theta_i - b_j)$$

其中, Y_{ij} 表示学习者*i*对问题*j*的回答; θ_i 表示学习者*i*的能力; a_j 表示问题*j*的区分度; b_j 表示问题*j*的难度; c_j 表示猜对概率; d_j 表示失误概率。

当 $d_j = 1$ 时得到3PLM-IRT;当 $d_j = 1, c_j = 0$ 时得到2PLM-IRT;当 $d_j = 1, c_j = 0, a_j = 1$ 时得到1PLM-IRT(Rasch模型),如图2所示。1PLM-IRT模型如式(7)所示,可以看到,1PLM-IRT将Logistic函数的指数部分扩展为学习者能力和题目难度的线性函数 z' 。

$$P(Y_{ij}=1|\theta_i, b_j) = \frac{1}{1 + e^{-z'}} \quad (7)$$

$$z' = \theta_i - b_j$$

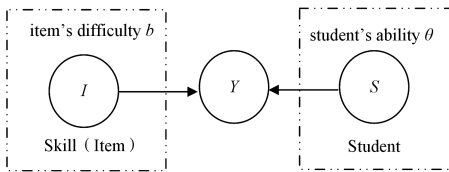


图2 1PLM-IRT(Rasch)模型图

Fig. 2 Diagram of 1PLM-IRT(Rasch)model

IRT模型把学习者的能力视为静态和一维的参数,适用于学习者能力不变的环境,例如以考试为代表的总结性评价。然而,在学习过程中,学习者的能力不是静态的,而是动态变化的,因此IRT不适合对学习者的能力存在变化的学习过程进行建模。

为了解决学习过程中学习者能力的时变性特征,卡耐基梅隆大学的Koedinger^[4]研究团队提出了AFM,如图3所示。通过对比图2和图3可以看出,AFM在Rasch模型的基础上进一步扩展了参数,考虑了学习率以及练习次数这两个变量

对学习者的认知状态的影响。AFM的具体模型如式(8)所示:

$$P(Y_{ij}=1) = \frac{1}{1 + e^{-z_{ij}}} \quad (8)$$

$$z_{ij} = \theta_i + \sum_k q_{jk}(\beta_k + \gamma_k T_{ik})$$

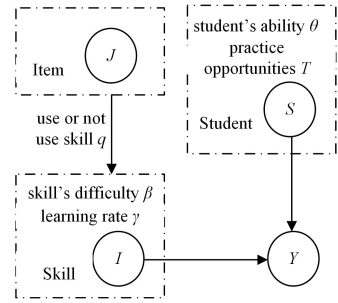


图3 AFM模型图

Fig. 3 Diagram of AFM model

其中,学习者参数 θ_i 和 T_{ik} 分别表示学习者*i*解决问题*j*所具有的先验知识(能力)和学习者*i*在第*k*个知识点上的练习次数。问题参数包括 β_k, γ_k 和 q_{jk} 。其中, β_k 表示问题*j*中第*k*个知识点的难度, γ_k 表示问题*j*中第*k*个知识点的学习率; q_{jk} 表示问题*j*中是否包含第*k*个知识点, q_{jk} 为1表示问题*j*中包含第*k*个知识点, q_{jk} 为0则相反。假设学习者需要完成*m*道含有相同知识点的题目,已知题目的 Q 矩阵(问题知识点对应矩阵)、学生前*m*-1道题目正确与否以及每一题尝试次数的情况下,用最大似然估计法训练得到 θ, β, γ 参数,最后预测学生在第*m*题上能否回答正确。与BKT不同,AFM假设学习是一个渐进的变化过程而不是离散的过渡,不是估计学习者潜在的知识状态,而是直接预测学习者回答正确的概率。首先,不同学习者的基础知识掌握能力不同,需要设置能力参数(能力不再是静态不变的)。其次,题目可以是多知识点的。再次,学习率只反映知识点是否容易学,与学习者个体无关(学习率是关于知识点的参数);每个知识点的难度不一样,需要给每个知识点设置难度参数和学习率。可以看出,AFM所强调的能力渐变性和多知识点耦合更符合真实的学习情境,因此获得了广泛的应用。

3.2 模型改进

从IRT到AFM,逻辑模型的适用范围从原来的总结性评价扩大到过程性评价,实现了从静态到动态的评估跨越。理论上,多知识点问题比单知识点问题难度大。为了修正难度系数,Cen等对AFM的参数进行了约束,得出了结合因素模型(Conjunctive Factor Model, CFM)^[4,51]。Pavlik等^[52]在AFM的基础上再进行了扩展,将练习次数细分为做对次数和做错次数,形成了成绩因素分析(Performance Factor Analysis, PFA)。与BKT相比,AFM和PFA都不具有失误参数,也就是说,学习者随着练习机会的增多,回答正确的概率一定是单调递增的,而不考虑其他因素可能导致的下降。MacLellan等^[53]进行了新的逻辑回归修正,该修正增加了失误参数,用于解释现实生活中切实存在的由于学习者自身原因导致的一类消极情况,例如掌握了知识点却失误。修正后的模型AFM+Slip与PFA+Slip比AFM和PFA更好地解释了因为失误等行为导致的情况,也具有更高的预测精度^[54]。

BKT 和 AFM 都属于一类高度结构化的模型,其参数都具有教学上的可解释性。但是,BKT 需要依赖专家标注领域知识,而 AFM 中使用的 Q 矩阵需要建立知识点与问题之间的映射,同样需要专家的参与。DKT 的出现很好地解决了这类问题,降低了人工成本。

4 深度知识追踪 (DKT)

4.1 模型原理

对学习者的知识状态建模的主要困难在于学习者个体和知识的内在复杂性。人类的学习过程本身就是抽象的思维过程,而知识更不是独立的存在。与隐马尔可夫模型相比,循环神经网络 (Recurrent Neural Networks, RNN) 在表征状态时具有高维、连续的特征,更适合于复杂模型的建立。并且,

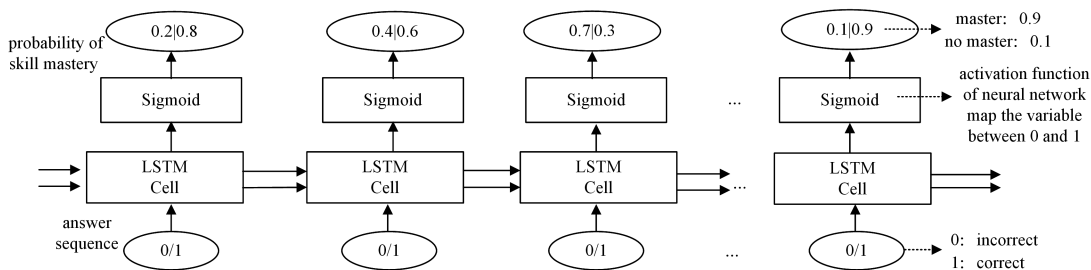


图 4 深度知识追踪模型图

Fig. 4 Diagram of deep knowledge tracing model

4.2 模型改进

DKT 只考虑了问题的知识点组成,将学习者回答的正确性作为输入,并未应用其他的数据信息。Zhang 等^[55]在 DKT 中引入更多学习者和问题的特征以扩展 DKT,包括正确率、学习者响应时间、尝试次数、第一次响应结果等。Mongkhonvanit 等^[56]在 DKT 中加入课程交互协变量来考虑学习者的参与度。Minn 等^[57]先在一定时间间隔内动态地将学生分成具有相似能力的不同群体,再把分类信息整合到 DKT 中。Nagatani 等^[58]通过合并与遗忘相关的多种类型的信息来考虑遗忘行为。Yeung 等^[59]在 DKT 的基础上增加了从学生档案中提取的其他特征,从预测学习者成绩扩展为预测其大学毕业后的第一份工作是否属于 STEM 领域。

DKT 不能准确地确定学习者对于知识概念的掌握程度。Zhang 等^[60]提出了一种新的动态键值存储网络模型 (Dynamic Key-Value Memory Networks, DKVMN),模型有一个静态矩阵 key 和一个动态矩阵 value,分别存储和更新相应概念

其变体长短期记忆 (Long Short-Term Memory, LSTM) 神经网络能够更充分地利用时间序列信息,在处理许多时间序列问题上具有优越性。Chris 等^[5]提出了用深度学习来处理知识追踪,形成了 DKT 模型。DKT 通过运用 LSTM 来预测知识状态,输入学习者的历史成绩,通过独热 (one-hot) 或压缩感知算法将输入的数据压缩转化为向量表示,经输入层提取特征,并将此特征输入隐藏层,再从输出层输出预测结果,即学习者下一次答题的成绩,如图 4 所示。DKT 与 BKT 相比具有 4 个优势^[22]: 1) DKT 能更多地利用学习者最近的表现来预测成绩; 2) DKT 针对多个知识点建模; 3) 神经网络可以依据学习者的成绩数据,得出知识点之间的关系; 4) DKT 能根据学生在各个知识点上的成绩得出学习者个体之间的差异。

的掌握程度。该模型不仅可以发现底层概念之间的关系,还可以直接输出学生对每个概念的掌握程度。Ha 等^[61]在 DKVMN 的基础上改进了学习遗忘机制。Abdelrahman 等^[62]建立了序列键值记忆网络 (Sequential Key-Value Memory Networks, SKVMN),该模型结合了 DKT 模型的循环建模能力和 DKVMN 模型的记忆能力,提高了预测精度。DKT 作为一类高度复杂的通用模型,其参数不具有可解释性。Yeung^[63]综合了 IRT 模型和 DKVMN 模型,使基于深度学习的知识跟踪具有一定的可解释性。当然,DKT 也可以用于一些开放式的学习环境中,如编程练习等^[64],为学习者的学习行为提供有价值的见解。

综上,BKT 模型、AFM 模型和 DKT 模型这 3 类主流的知识追踪模型都是通过历史成绩数据来预测学习者的知识掌握状态或答题正确率,并在基本模型的基础上纳入不同的数据信息,以更好地反映学习者的学习过程,提高预测精度。BKT, AFM, DKT 又具有一定的差异,如表 1 所列。

表 1 模型主要特点的比较

Table 1 Comparison of main features of three models

模型	原理	关于学习的假设	优点	局限性	代表研究
BKT	隐马尔可夫模型	离散过渡	模型简单 具有教学解释性	依赖专家标注 针对单个知识点建模	增加时间因素 ^[21-22] 个性化研究 ^[27-28,30]
AFM	逻辑回归模型	渐进变化	模型简单 增加 Q 矩阵 具有教学解释性	依赖专家标注	PFA ^[52] AFM+ ^[53]
DKT	循环神经网络	无假设	性能较优 解决了对专家的 依赖问题	模型复杂 不具教学解释性 训练数据量庞大	增加多源数据 ^[55-56] DKVMN ^[60] SKVMN ^[62]

BKT 在建模过程中假设学习是知识状态从未知到已知

的离散过渡; AFM 假设学习是一个渐进的变化过程; DKT 则

不涉及学习过程的假设。BKT 只适用于单个知识点,而 AFM 与 DKT 适用于多知识点。BKT 和 AFM 都是高度结构化的模型,其参数都具有教学的解释性。DKT 不具有教学的解释性,虽然预测精度高且不依赖于专家的标注,但依赖于庞大的数据。

另外,有一类研究开始探讨领域模型中所关注的知识点对模型的重要性。Montero 等^[65]通过实验比较了 DKT 和 BKT,认为 DKT 多知识点的交叉输入是 DKT 获得较好预测结果的原因之一。Hawkins 等^[66]和 Wang 等^[67]的研究也提到了知识点对应的问题之间相似性的重要性。通过梳理文献发现,领域模型中知识点之间的关系建模的研究与知识空间理论的内容具有天然的相似性,因此笔者接下来对知识空间理论进行简要介绍,并以此为出发点提出研究展望。

5 知识空间理论(KST)

5.1 理论起源

KST 将人类的认知理解用数学的方式表达出来,并且形成了严谨的数学理论,也被认为是一种教育测量理论^[68]。该理论主要应用于学习辅助及自适应测评领域^[69]。KST 认为,在传统评估方式(考试成绩或排名等)下,学习者只知道自己在哪一分数层或类别,这对他们明确自己处于怎样的知识状态水平以及接下来要如何学习没有明确的帮助。这样的评估显得过于粗糙。

KST 基于知识点组合及概率模型,形成知识空间,并对学习者所处的学习情况,即知识状态进行评估。KST 中的相关定理结合可实现的技术操作,能够准确地定位学习者当前所处的知识状态,由此为学习者提供进一步的指导^[70]。KST 与一般的数值评价的区别在于,它不仅考虑了学习者回答的正确性,还揭示了个人的知识状态,产生的评估结果比简单的数值评价更具有指导性。需要注意的是,KST 揭示的知识状态与 BKT 有所不同。BKT 预测的是学习者是否掌握当前知识点,KST 所揭示的是学习者所掌握的知识点的精确集合。

5.2 理论内容

知识空间理论中相关定理对知识追踪模型的构建具有较大的价值。本文首先介绍知识空间的相关术语。项目(item)表示一种问题类型,常用 q 表示;领域(domain)表示由项目所组成的庞大的问题集合,常用 Q 表示;知识状态(Knowledge State)是 Q 的子集,常用 K 和 L 表示。

定义 1^[69] 我们称 (Q, \mathcal{K}) 为知识结构(knowledge structure), (Q, \mathcal{K}) 包含非空集合 Q 和 Q 的子集的集合 \mathcal{K} , 且 $\emptyset \in \mathcal{K}, Q \in \mathcal{K}$ 。当 Q 有限时, (Q, \mathcal{K}) 也有限。一般将 (Q, \mathcal{K}) 简写为 \mathcal{K} 。

注意, \mathcal{K} 与 K 均为 Q 的子集。但 \mathcal{K} 表示包含多种状态的状态集合,而 K 常常表示具体的某个知识状态,两者在包含的范围空间上有大小差异。

定义 2^[69] 满足以下两个准则的知识结构称为学习空间(Learning Space)。

(1)学习平滑性: $\forall K, L \in \mathcal{K}, K \subset L$, 存在有限状态链 $K = K_0 \subset K_1 \subset \dots \subset K_p = L$, 其中 $|K_i \setminus K_{i-1}| = 1, 1 \leq i \leq p$ ($|L \setminus K| = p$)。

(2)学习一致性: $\forall K, L \in \mathcal{K}, K \subset L$, 如果 $K \cup \{q\}$ 是一个知识状态,那么 $L \cup \{q\}$ 也是一个知识状态。

定义 3^[69] 我们称对并运算封闭的知识结构为知识空间(Knowledge Space),即对于 $\forall K, L \in \mathcal{K}$, 有 $K \cup L \in \mathcal{K}$ 。

边缘定理^[69] 假设 (Q, \mathcal{K}) 为知识结构,则:

在 \mathcal{K} 中某个 K 的内边缘为满足以下条件的项的集合:
 $K^I = \{q \in K \mid K \setminus \{q\} \in \mathcal{K}\}$;

在 \mathcal{K} 中某个 K 的外边缘为满足以下条件的项的集合:
 $K^O = \{q \in Q \setminus K \mid K \cup \{q\} \in \mathcal{K}\}$ 。

注意,全集 Q 有一个空的外边缘,空集 \emptyset 有一个空的内边缘。

严格的知识空间规定,任何知识状态都在一条“学习路径”上。除初始状态外,路径上的任何一个状态都比前面状态多一个项目,这样的一条学习路径称为一个层次(graduation)学习路径^[71]。如果知识空间中的任何一个状态都存在于一条层次学习路径上,则这个知识空间是级配良好的(well-graded)^[72]。学习空间在理论上等价于级配良好的知识空间。

知识状态的内外边缘具有重要的教学意义。外边缘表示学生已经准备好去学习的知识(ready-to-learn knowledge),内边缘表示学生能力的高点(high points of competence)^[73]。在一个学习空间中,任何状态都可由它的内外边缘指定。如果学习空间是某个领域的知识或规则的准确表征,那么外边缘就能准确地告诉我们学生接下来准备学什么^[74]。这给个性化教学带来了巨大的价值。

5.3 融合知识点先决关系的可加因素模型

知识空间中知识点的关系侧重于知识点之间的先决关系。知识空间的学习路径及知识状态的内外边缘表明,将知识点之间的先决关系纳入学习者模型中,在一定程度上不仅可以提高预测精度,还可以更好地为学习者推荐学习资源。本文从数学模型的角度,将知识点之间的先决关系与知识追踪技术联系起来,提出了一种融合知识点先决关系的可加因素模型。该方法通过将学习项目的 Q 矩阵与知识点的先决关系矩阵(P 矩阵)相乘,得到一个新的矩阵(pq -矩阵),该矩阵不仅可以反映出一个学习项目中包含的知识点,还能体现出知识点之间的先决关系。进一步,将 pq -矩阵嵌入到 AFM 模型中进行知识追踪。限于篇幅,我们在此通过一个案例来说明 pq -矩阵的生成方法。例如,某自适应学习系统为学习者推送了由 3 道题目组成的知识点练习任务,该任务包含 4 个相关知识点(A, B, C, D)。假设该任务的 Q 矩阵(3×4 矩阵)如下:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

其中,第一道题目包含 A 和 C 知识点,第二道题目包含 A, B, D 知识点,第三道题目包含 D 知识点。在知识点的先决关系上,我们假设 A 是 B 的先序, B 是 D 的先序, C 为独立知识点,则可得到这 3 道题目的先决关系矩阵 P 矩阵为:

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

矩阵中的元素 p_{ij} 表示知识点是否具有先决关系。若 p_{ij} 为 1, 则 i 是 j 的先序, 反之则不具有先决关系。注意, 该矩阵设定知识点本身与自己存在关系, 即对角线元素为 1。通过将 Q 矩阵乘以 P 矩阵的转置得到 pq -矩阵。计算过程如下:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \rightarrow \begin{bmatrix} 1 & 0 & 1 & 0 \\ 3 & 2 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

可以看到, pq -矩阵中的每一项表示某个知识点在该任务中被练习的次数。例如, 题目二包含 3 个知识点(A、B、D), 而在结构上 A 是 B 的先序, B 是 D 的先序, 因此对于知识点 A, 存在 1 次直接考查, 2 次间接查。即该题对于知识点 A 的考查出现了 3 次, 学习者的练习次数 (opportunities) 为 3, 而在传统的 AFM 中知识点的练习次数仅为 1 次, 可以看出改进后的模型更符合逻辑。

6 研究结论

如前文所述, 3 类主流的知识追踪模型在理论研究与应用研究上都取得了一定的成功。然而, 以上研究也存在一定的不足。3 类模型较少涉及知识点之间关系的建模, 或者建模方式太过简略。值得一提的是, Carmona 等^[75] 在多层贝叶斯学生模型中就已经开始研究知识点之间的先决关系对预测的影响。Chen 等^[76] 将具备先决关系的知识点概念建模为有序对 (ordering pair) 作为 DKT 的约束条件。Meng 等^[77] 将知识点之间的先决关系以矩阵的形式加入到 BKT 中。Xu 等^[78] 使用逻辑回归方法扩展 BKT 参数, 提出了 3 种面向多知识点的模型。这些都表明, KST 所关注的知识模型, 尤其是将知识点之间的先决关系加入到知识追踪模型中, 可以在一定程度上提高预测的准确性, 帮助我们更好地了解学习者的知识状态, 提供更合适的指导, 这也是下一阶段的研究重点。

另外, DKT 在预测上虽然比其他模型更好, 但由于它所需要的训练数据量大, 模型本身比较复杂, 也不具有良好的教学可解释性, 因此并不足以取代 BKT 和 AFM 模型^[79-80]。笔者建议, 实践者在自适应学习系统的初始应用阶段可采用 AFM 模型来建立学习者模型, 当系统积累了大量的数据之后, 可考虑嵌入 DKT 模型, 或并行使用两种模型。笔者认为, 将知识点之间的结构更好地加入到这两个模型中以提升预测准确性是下一阶段研究的重点方向, 如图 5 所示。

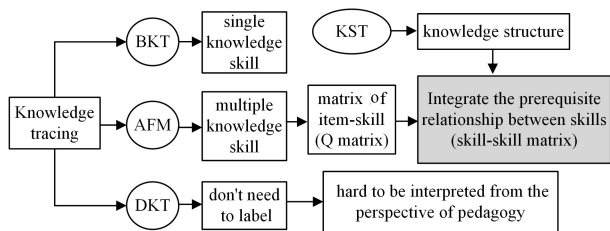


图 5 融合知识点关系的知识追踪

Fig. 5 Knowledge tracking integrating knowledge skill relationship of skills

矩阵中添加知识点先决关系矩阵, 下一步计划将改进后的模型应用于真实学习数据, 通过分析学习曲线的差异, 探究知识点关系是否对数据的预测产生影响, 进而分析能否改善在线学习系统和提升学习者的学习效率。

结束语 本文对现有的 3 类知识追踪技术进行了综述。通过对国内外相关文献的分析得出结论: 贝叶斯知识追踪的研究与应用最早也最为广泛。但由于其对单一知识点建模, 因此对于从知识点层面进行多维度建模具有局限性。可加因素模型融合了题目与知识点之间的关系, 增加了建模维度, 但与贝叶斯知识追踪模型相同的是, 两者都需要领域专家的参与, 增加了时间成本。深度知识追踪的出现很好地解决了这一问题, 但随之而来的问题包括对数据量和计算量的要求等。近年来, 关于知识追踪的新的研究逐渐增加, 但研究与应用过程还需要关注很多问题。本文从知识空间的角度出发, 认为在研究过程中应该增加知识点先决关系的建模以改进现有的模型。当然, 未来的研究不仅要在数据建模层面增加建模的维度, 还应该从适用的环境出发, 结合不同领域进行建模, 与此同时, 深度学习和大数据的发展应该成为知识追踪技术改革和创新的关键。本文从以上分析得出知识追踪未来的主要研究发展方向如下:

(1) 增加建模的维度。现有的研究绝大多数都是基于学习者与智能教辅系统的交互数据建立模型进行预测, 这类数据通常是历史学习成绩, 建模的维度比较单一。本文提出从知识点层面增加建模维度, 即在原有模型中融入知识点关系矩阵。此外, 还可以从学习者层面进行研究。学习情绪是目前较为关注的研究方向, 通常包括困惑、投入、沮丧、厌倦等。在不同的学习者身上, 相同的情绪唤醒程度具有高低差异; 不同的学习情绪对学习本身会造成消极或积极的影响, 从而影响学习结果。因此, 在未来智能教辅系统的研究中可以更加关注学习者在学习过程中产生的学习情绪及其变化, 及时提供反馈并在适当的时候给予教学干预, 帮助学习者远离负面情绪, 保持积极的学习状态。另一方面, 对于学习者特征, 如何选择合适的情绪作为建模因子、如何将情绪与其他数据融合进行预测、如何提高模型性能等, 都是未来的研究方向。

(2) 针对不同领域建模。知识追踪模型的选择取决于建模的具体背景。数学学科知识点概念清晰, 知识追踪模型一开始主要针对数学学科建立。由于不同学科具有不同的特征, 针对单一学科建立的模型可移植性不佳。例如开放式的学习环境, 包括编程学习和语言学习。编程学习又可以分为文本式和图形化编程, 输入的数据通常为代码或者相应的代码图形。语言学习输入的数据又包含文本和语音。因此, 未来的研究应该根据不同学科的特点, 尝试使用其他算法或引入其他模型, 使得知识追踪的适用环境更具针对性。

(3) 深度学习与大数据技术的应用。一方面, 数字化学习平台存储了大量的学习者学习数据, 而深度学习的发展提高了大数据的利用率。另一方面, 高度依赖于专家标注的模型在处理异构数据上具有局限性。深度学习在处理语音信息、图像识别等方面表现出较好的性能。在此趋势下, 知识追踪模型的构建可以更好地依赖于客观数据。因此, 未来的研究可以着重于深度知识追踪并探索如何解决不具教学解释性的问题。

参 考 文 献

- [1] ZHANG G, ZHOU D D, GE Q Q. A Review of Learner Feature Models and Modeling Methods in Adaptive Learning Systems [J]. *Modern Educational Technology*, 2012, 22(5): 77-82.
- [2] YUDELSON M V, KOEDINGER K R. Estimating the benefits of student model improvements on a substantive scale[C]// *Proceedings of the 6th International Conference on Educational Data Mining, EDM*, Memphis, USA: 2013.
- [3] CORBETT A T, ANDERSON J R. Knowledge tracing: Modeling the acquisition of procedural knowledge[J]. *User Modeling and User-Adapted Interaction*, 1995, 4(4): 253-278.
- [4] CEN H. Generalized Learning Factors Analysis: Improving cognitive Models with Machine Learning[D]. Pittsburgh: Carnegie Mellon University, 2009.
- [5] PIECH C, SPENCER J, HUANG J, et al. Deep Knowledge Tracing[C]// *Advances in Neural Information Processing Systems*. 2015: 505-513.
- [6] XU P F, ZHENG Q H, CHEN Y H, et al. Learner modeling analysis in Educational Data Mining[J]. *Distance Education in China*, 2018(6): 5-12.
- [7] DESMARAIS M C, BAKER R S J D. A review of recent advances in learner and skill modeling in intelligent learning environments[J]. *User Modeling and User-Adapted Interaction*, 2012, 22(1/2): 9-38.
- [8] PELÁNEK R. Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques[J]. *User Modeling and User-Adapted Interaction*, 2017, 27(3-5): 313-350.
- [9] LI F M, YE Y W, LI X F, et al. Application of knowledge tracing Model in education: a review of relevant studies from 2008 to 2017 [J]. *Distance Education in China*, 2019(7): 86-91.
- [10] LIU H Y, ZHANG T C, WU P W, et al. A Review of Knowledge Tracing [J]. *Journal of East China Normal University (Natural Science)*, 2019(5): 1-15.
- [11] FALMAGNE J C, DOIGNON J P. Learning spaces: Interdisciplinary applied mathematics[M]. Springer Science & Business Media, 2011.
- [12] YE Y W, LI F M, LIU Q Q, et al. Incorporating the variables of forgetting and data volume into Knowledge Tracing Model: how does it impact prediction accuracy? [J]. *Distance Education in China*, 2019(8): 20-26.
- [13] BLOOM B. Learning for Mastery [J]. *Evaluation Quarterly*, 1968, 1(2): 1-12.
- [14] BOZHOVICH E D. Zone of Proximal Development[J]. *Journal of Russian & East European Psychology*, 2009, 47(6): 48-69.
- [15] LI H. Statistical learning method[M]. Beijing: Tsinghua University Press, 2012: 171-174.
- [16] ZHANG K, YAO Y. A three learning states Bayesian knowledge tracing model [J]. *Knowledge-Based Systems*, 2018, 148: 189-201.
- [17] AGARWAL D, BAKER R. Dynamic knowledge tracing through data driven recency weights[C]// *Proceedings of the 27th International Conference on Computers in Education*. 2019.
- [18] WANG Y, HEFFERNAN N T, BECK J E. Representing Student Performance with Partial Credit[C]// *Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining*, Pittsburgh, PA, USA, 2010: 11-13.
- [19] WANG Y, HEFFERNAN N. Extending knowledge tracing to allow partial credit: Using continuous versus binary nodes[C]// *International Conference on Artificial Intelligence in Education*. Berlin, Heidelberg: Springer, 2013: 181-188.
- [20] PARDOS Z A, HEFFERNAN N T, RUIZ C, et al. Effective Skill Assessment Using Expectation Maximization in a Multi Network Temporal Bayesian Network[C]// *In Proceedings of the The Young Researchers Track at the 9th International Conference on Intelligent Tutoring Systems*. 2008.
- [21] QIU Y, QI Y, LU H, et al. Does time matter? Modeling the effect of time with bayesian knowledge tracing[C]// *Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011)*. 2011 (Table D): 139-148.
- [22] KHAJAH M, LINDSEY R V, MOZER M C. How deep is knowledge tracing? [C]// *Proceedings of the 9th International Conference on Educational Data Mining, EDM*, 2016: 94-101.
- [23] LYU H Y, SHEN L, QI M. Knowledge tracing model and Integration Technology based on "attitude" [J]. *Journal of Jiangsu Normal University (Natural Science Edition)*, 2011, 29(4): 54-57.
- [24] ZHU J, ZANG Y, QIU H, et al. Integrating Temporal Information into Knowledge Tracing: A Temporal Difference Approach [J]. *IEEE Access, IEEE*, 2018, 6: 27302-27312.
- [25] CEN H, KOEDINGER K R, JUNKER B. Is over practice necessary? Improving learning efficiency with the Cognitive Tutor through educational data mining[C]// *Proceedings of the 13th International Conference on Artificial Intelligence in Education AIED 2007*. 2007: 511-518.
- [26] LEE J I, BRUNSKILL E. The Impact on Individualizing Student Models on Necessary Practice Opportunities[C]// *International Educational Data Mining Society, Paper presented at the International Conference on Educational Data Mining, EDM*, Chania, Greece, 2012.
- [27] BAKER R S, CORBETT A T, ALEVEN V. More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing[C]// *International Conference on Intelligent Tutoring Systems*. Berlin, Heidelberg: Springer, 2008: 406-415.
- [28] PARDOS Z A, HEFFERNAN N T. Modeling individualization in a Bayesian networks implementation of knowledge tracing [C]// *International Conference on User Modeling, Adaptation, and Personalization*. Berlin, Heidelberg: Springer. 2010: 255-266.
- [29] NEDUNGADI P, REMYA M S. Predicting students' performance on intelligent tutoring system-Personalized clustered BKT(PC-BKT) model[C]// *Frontiers in Education Conference (FIE) Proceedings. IEEE*, 2014: 1-6.
- [30] YUDELSON M V, KOEDINGER K R, GORDON G J. Individualized bayesian knowledge tracing models[C]// *International Conference on Artificial Intelligence in Education*. Berlin, Hei-

- delberg; Springer, 2013; 171-180.
- [31] PARDOS Z A, HEFFERNAN N T. KT-IDEM: introducing item difficulty to the knowledge tracing model[J]. *Lecture Notes in Computer Science*, 2011, 6787: 243-254.
- [32] XU Y, CHANG K. Using EEG in Knowledge Tracing[C]// *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, 2014; 361-362.
- [33] SPAULDING S, BREAZEL C. Affect and Inference in Bayesian Knowledge Tracing with a Robot Tutor[C]// *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 2015; 219-220.
- [34] LIN C, CHI M. Intervention-BKT: Incorporating Instructional Interventions into Bayesian Knowledge Tracing[C]// *International Conference on Intelligent Tutoring Systems*. Cham: Springer, 2016; 208-218.
- [35] GONZÁLEZ-BRENES J, HUANG Y, BRUSILOVSKY P. General features in knowledge tracing: Applications to multiple subskills, temporal item response theory, and expert knowledge [C]// *The 7th International Conference on Educational Data Mining*, University of Pittsburgh, 2014; 84-91.
- [36] BECK J E, CHANG K M, MOSTOW J, et al. Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology[C]// *International Conference on Intelligent Tutoring Systems*. Berlin, Heidelberg: Springer, 2008; 383-394.
- [37] PARDOS Z, BERGNER Y, SEATON D, et al. Adapting Bayesian Knowledge Tracing to a Massive Open Online Course in edX [C]// *Proceedings of the 6th International Conference on Educational Data Mining*. Memphis, 2013.
- [38] WANG Z, ZHANG M. Evaluation of MOOC students based on Bayesian Knowledge Tracing Model [J]. *China Science Paper*, 2015, 10(2): 241-246.
- [39] WANG Z, ZHU J, LI X, et al. Structured Knowledge Tracing Models for Student Assessment on Coursera[C]// *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*. ACM, 2016; 209-121.
- [40] PARDOS Z A, GOWDA S M, BAKER R S J D, et al. Ensembling predictions of student post-test scores for an intelligent tutoring system[C]// *Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011)*, 2011; 189-198.
- [41] TARIQ Q, KOLCHINSKI A, DAVIS R. Modelling Student Knowledge as a Latent Variable in Intelligent Tutoring Systems: A Comparison of Multiple Approaches[EB/OL]. [2012-12-27]. <http://cs229.stanford.edu/proj2016/poster/Tariq-Kolchinski-Davis-ModelingStudentKnowledgeAsALatentVariable-poster.pdf>.
- [42] DEONOVIC B, YUDELSON M, BOLSINOVA M, et al. Learning meets Assessment: On the relation between Item Response Theory and Bayesian Knowledge Tracing[J]. *Behaviormetrika*, 2018, 45(2): 457-474.
- [43] KHAJAH M M, HUANG Y, GONZÁLEZ-BRENES J P, et al. Integrating knowledge tracing and item response theory: A tale of two frameworks[C]// *CEUR Workshop Proceedings*, 2014; 7-15.
- [44] CUI Y, CHU M W, CHEN F. Analyzing Student Process Data in Game-Based Assessments with Bayesian Knowledge Tracing and Dynamic Bayesian Networks[J]. *Journal of Educational Data Mining*, 2019, 11(1): 80-100.
- [45] THAI-NGHE N, BUSCHE A, SCHMIDT-THIEME L. Improving Academic Performance Prediction by Dealing with Class Imbalance[C]// *2009 9th International Conference on Intelligent Systems Design and Applications*, 2009; 878-883.
- [46] RITTER S, HARRIS T K, NIXON T, et al. Reducing the Knowledge Tracing Space[C]// *Proceedings of the 2nd International Conference on Educational Data Mining*. DBLP, Cordoba, Spain, 2009.
- [47] FANCSALI S, NIXON T, RITTER S. Optimal and Worst-Case Performance of Mastery Learning Assessment with Bayesian Knowledge Tracing[C]// *Educational Data Mining*, 2013.
- [48] DHANANI A, LEE S Y, PHOTHILIMTHANA P M, et al. A comparison of error metrics for learning model parameters in Bayesian knowledge tracing [C] // *CEUR Workshop Proceedings*, 2014, 1183; 153-154.
- [49] HAWKINS W J, HEFFERNAN N T, BAKER R S J D. Learning bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities[C]// *International Conference on Intelligent Tutoring Systems*. Cham; Springer, 2014; 150-155.
- [50] SLATER S S, BAKER R. Degree of Error in Bayesian Knowledge Tracing Estimates From Differences in Sample Sizes[J]. *Behavior metrika*, 2018, 45(2): 475-493.
- [51] CEN H, KOEDINGER K, JUNKER B. Comparing Two IRT models for conjunctive skills[C]// *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, 5091 LNCS (1): 796-798.
- [52] PAVLIK P I, CEN H, KOEDINGER K R. Performance factors analysis - A new alternative to knowledge tracing[J]. *Frontiers in Artificial Intelligence and Applications*, 2009, 200(1): 531-538.
- [53] MACLELLAN C J, LIU R, KOEDINGER K R. Accounting for Slipping and Other False Negatives in Logistic Models of Student Learning [C] // *Proceedings of the 8th International Conference on Educational Data Mining*, 2015; 53-60.
- [54] MACLELLAN C J. Investigating the impact of slipping parameters on additive factors model parameter estimates[C]// *CEUR Workshop Proceedings*, 2016; 1633.
- [55] ZHANG L, XIONG X, ZHAO S, et al. Incorporating Rich Features into Deep Knowledge Tracing [C] // *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. ACM, 2017; 169-172.
- [56] MONGKHONVANIT K, KANOPKA K, LANG D. Deep knowledge tracing and engagement with MOOCs[C]// *ACM International Conference Proceeding Series*, 2019; 340-342.
- [57] MINN S, YU Y, DESMARAIS M C, et al. Deep Knowledge Tracing and Dynamic Student Classification for Knowledge Tracing[C]// *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2018; 1182-1187.
- [58] NAGATANI K, CHEN Y Y, ZHANG Q, et al. Augmenting

- knowledge tracing by considering forgetting behavior[C]// the World Wide Web Conference. 2019;3101-3107.
- [59] YEUNG C K, YEUNG D Y. Incorporating Features Learned by an Enhanced Deep Knowledge Tracing Model for STEM/Non-STEM Job Prediction[C]// International Journal of Artificial Intelligence in Education. 2018;1-25.
- [60] ZHANG J, SHI X, KING I, et al. Dynamic Key-Value Memory Networks for Knowledge Tracing[J]. Proceedings of the 26th International Conference on World Wide Web. 2017;765-774.
- [61] HA H, HWANG U, HONG Y, et al. Memory-Augmented Neural Networks for Knowledge Tracing from the Perspective of Learning and Forgetting[J]. arXiv:1805.10768, 2018.
- [62] ABDELRAHMAN G, WANG Q. Knowledge tracing with sequential key-value memory networks[C]// Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019). 2019; 175-184.
- [63] YEUNG C. Deep-IRT: Make Deep Learning Based Knowledge Tracing Explainable Using Item Response Theory[J]. arXiv: 1904.11738, 2019.
- [64] WANG L, SY A, LIU L, et al. Learning to represent student knowledge on programming exercises using deep learning[C]// Proceedings of the 10th International Conference on Educational Data Mining, EDM, 2017; 324-329.
- [65] MONTERO S, ARORA A, KELLY S, et al. Does deep knowledge tracing model interactions among skills? [C]// Proceedings of the 11th International Conference on Educational Data Mining, EDM, 2018; 462-466.
- [66] HAWKINS W J, HEFFERNAN N T. Using Similarity to the Previous Problem to Improve Bayesian Knowledge Tracing [C]// EDM(Workshops). 2014.
- [67] WANG Z, FENG X, TANG J, et al. Deep Knowledge Tracing with Side Information[C]// International Conference on Artificial Intelligence in Education. Cham; Springer International Publishing. 2019; 303-308.
- [68] FU Q, SUN B. Comparative study of Knowledge Space Theory and Item Response Theory [J]. China Educational Technology, 2014(5); 75-76.
- [69] DOIGNON J P, FALMAGNE J C. Knowledge spaces and learning spaces[C]// New Handbook of Mathematical Psychology. 2016; 274-321.
- [70] FALMAGNE J C, ALBERT D, DOBLE C, et al. Knowledge spaces: Applications in Education[M]. Springer Science & Business Media, 2013.
- [71] BARTL E, BELOHLAVEK R. Knowledge spaces with graded knowledge states[J]. Information Sciences, 2011, 181(8); 1426-1439.
- [72] VILLANO M. Probabilistic student models: Bayesian belief networks and knowledge space theory[C]// International Conference on Intelligent Tutoring Systems. Berlin, Heidelberg; Springer, 1992; 491-498.
- [73] DOBLE C, MATAYOSHI J, COSYN E, et al. A Data-Based Simulation Study of Reliability for an Adaptive Assessment Based on Knowledge Space Theory[J]. International Journal of Artificial Intelligence in Education, 2019, 29(2); 258-282.
- [74] FALMAGNE J C. Markov Processes in Learning Spaces[J]. International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 2015, 14; 584-590.
- [75] CARMONA C, MILLÁN E, PÉREZ-DE-LA-CRUZ J L, et al. Introducing prerequisite relations in a multi-layered Bayesian student model[C]// International Conference on User Modeling. Berlin, Heidelberg; Springer, 2005; 347-356.
- [76] CHEN P, LU Y, ZHENG V W, et al. Prerequisite-Driven Deep Knowledge Tracing[C]// Proceedings - IEEE International Conference on Data Mining, ICDM, 2018; 39-48.
- [77] MENG L, ZHANG M, ZHANG W, et al. CS-BKT: introducing item relationship to the Bayesian knowledge tracing model [J/OL]. Interactive Learning Environments, [2020-12-27]. <https://www.tandfonline.com/doi/abs/10.1080/10494820.2019.1629600>.
- [78] XU M K, WU W J, ZHOU X, et al. Research on multi knowledge skill knowledge tracking model and visualization [J]. e-Education Research, 2018(10); 53-59.
- [79] XIONG X, ZHAO S, VAN INWEGEN E G, et al. Going deeper with deep knowledge tracing[C]// Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016). 2016; 545-550.
- [80] DOROUDI S, BRUNSKILL E. Fairer but not fair enough on the equitability of knowledge tracing[C]// ACM International Conference Proceeding Series. 2019; 335-339.



ZHANG Nuan, born in 1995, postgraduate. Her main research interests include big data of education, personalized learning and knowledge tracing.



JIANG Bo, born in 1985, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include educational data mining, learning analytics and machine learning.