计算机科学
COMPUTER SCIENCE

基于速度与准确率权衡的深度认知诊断模型

程艳, 周子为, 马明宇, 林庆龙, 詹勇鑫, 万凌峰

引用本文

程艳, 周子为, 马明宇, 林庆龙, 詹勇鑫, 万凌峰. 基于速度与准确率权衡的深度认知诊断模型[J]. 计算机科学, 2024, 51(10): 170-177.

CHENG Yan, ZHOU Ziwei, MA Mingyu, LIN Qinglong, ZHAN Yongxin, WAN Lingfeng. Speed-Accuracy Tradeoff-based Deep Cognitive Diagnostic Model [J]. Computer Science, 2024, 51(10): 170-177.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240300121

基于速度与准确率权衡的深度认知诊断模型

程　艳[1,2,3]　周子为[2,3]　马明宇[2,3]　林庆龙[2,3]　詹勇鑫[2,3]　万凌峰[2,3]
1 江西师范大学软件学院　南昌 330022
2 江西师范大学计算机信息工程学院　南昌 330022
3 江西省智能信息处理与情感计算省重点实验室　南昌 330022

摘　要　智能教育中,认知诊断通过分析学习者的学习行为数据来理解学习者的认知状态。现有基于深度学习方法的认知诊断模型默认假设学习者在作答过程中有足够的作答时间来完全发挥知识掌握水平,未考虑学习者在作答过程中的作答速度与作答准确率之间的权衡策略对发挥知识掌握水平的影响。针对上述问题,提出了一种基于速度与准确率权衡的深度认知诊断模型,首先构建认知风格模糊集解释学习者的权衡策略,然后通过动态逻辑回归函数模拟学习者作答过程中的速度与准确率权衡关系,实现对学习者理论上能达到最高的知识掌握水平与实际作答中发挥出来的知识掌握水平的区分诊断。此外还引入了作答时间属性和题目类型属性,以更准确地表征认知诊断交互函数中的题目参数。大量实验表明,该模型相比同类最优模型在3个公开数据集上准确度分别提升2.58%,2.86%,5.18%,且能为预测结果提供作答时间层面的解释,具有一定的优越性。

关键词:智能教育;深度认知诊断;速度与准确率权衡;模糊集;逻辑回归函数
中图分类号　TP391

Speed-Accuracy Tradeoff-based Deep Cognitive Diagnostic Model

CHENG Yan[1,2,3],ZHOU Ziwei[2,3],MA Mingyu[2,3],LIN Qinglong[2,3],ZHAN Yongxin[2,3] and WAN Lingfeng[2,3]
1 School of Software,Jiangxi Normal University,Nanchang 330022,China
2 School of Computer Information Engineering,Jiangxi Normal University,Nanchang 330022,China
3 Provincial Key Laboratory of Intelligent Information Processing and Affective Computing of Jiangxi Province,Nanchang 330022,China

Abstract　In intelligent education,cognitive diagnosis analyzes students' learning behavior data to understand their cognitive state. Existing cognitive diagnostic models based on deep learning methods assume by default that students have enough reaction time to fully exert the level of knowledge mastery during the response process,and do not consider the impact of the trade-off strategy between the speed and accuracy of student's response during the response process on the exertion of the level of knowledge mastery. Aiming at the above problem,a deep cognitive diagnostic model based on speed-accuracy trade-off is proposed,which firstly constructs a cognitive style fuzzy set to explain the students' trade-off strategy,and then simulates the speed-accuracy trade-off relationship in the process of the learners' response through the dynamic logistic regression function,so as to realize the differentiated diagnosis of the students' theoretically highest level of knowledge mastery from the level of knowledge mastery they have played out in the actual response. In addition,the reaction time attribute and exercise type attribute are introduced to more accurately characterize the topic parameters in the cognitive diagnostic interaction function. Numerous experiments show that the model not only improves the accuracy by 2.58%,2.86%,and 5.18% compared to similar optimal models on the three publicly available datasets,but also provides a superior explanation of the prediction results at the level of response time.

Keywords　Intelligent education,Deep cognitive diagnostics,Speed-Accuracy trade-off,Fuzzy sets,Logistic regression function

到稿日期:2024-03-18　返修日期:2024-07-07
基金项目:国家自然科学基金(62167006);江西省科技创新基地计划项目——智能信息处理与情感计算江西省重点实验室(2024SSY03131);江西省主要学科学术和技术带头人培养计划——领军人才项目(20213BCJL22047);江西省自然科学基金(20212BAB202017)
This work was supported by the National Natural Science Foundation of China(62167006),Jiangxi Provincial Science and Technology Innovation Base Program Project—Jiangxi Provincial Key Laboratory of Intelligent Information Processing and Affective Computing(2024SSY03131),Jiangxi Province Leading Talent Project of Major Academic Disciplines and Technologies(20213BCJL22047) and Natural Science Foundation of Jiangxi Province,China(20212BAB202017).
通信作者:程艳(chyan88888@jxnu.edu.cn)

1　引言

在数字化转型的浪潮中，人工智能等新兴信息技术的高速发展极大地推动了由数据驱动的教育变革[1]。人工智能(AI)技术与教育的深度融合是这场变革的核心，大数据和 AI 技术为学习者提供了更加个性化的学习体验，智能教育的出现改变了理解和实践教育的方式。认知诊断是实现这一目标的重要任务，其通过分析学习者的学习行为数据来理解学习者的认知状态[2]。

认知诊断来源于认知心理学，被广泛应用于教育测量任务，通过建立测量模型模拟学习者知识掌握水平与作答准确率之间的关系。传统认知诊断模型的特点是通过专家设计的学习者-题目交互函数对学习者进行认知评估时，同时对学习者参数与题目参数进行估计。其中最经典的传统认知诊断模型是项目反应理论(Item Response Theory，IRT)模型[3]与确定性输入噪声与门(Deterministic Input，Noisy and Gate，DINA)模型[4]，IRT 关注连续的学习者认知水平参数，而 DINA 更关注离散的学习者认知水平参数。学者们在两种经典模型上也进行了大量改进，例如，相对于 IRT，多维项目反应理论(Multidimensional Item Response Theory，MIRT)模型[5]在潜在特质的维度上进行扩展；广义确定性输入噪声与门(Generalized Deterministic Input，Noisy and Gate，G-DINA)模型[6]是 DINA 的一般化模型，可以处理更广泛类型的数据。这些传统认知诊断模型专注于对学习者-题目交互的模拟，但需要具有专业知识的专家进行人工设计。

随着智能教育的快速发展，在线学习平台的广泛应用带来了大量教育数据供学者们挖掘分析，机器学习等人工智能技术逐渐被应用于深度认知诊断(Deep Cognitive Diagnosis，DCD)建模中[7]。DCD 通过分析学习者的学习行为数据来理解学习者的认知状态，显化学习者认知水平，帮助学习者实现个性化学习，从而推动教育高质量发展。IRT 作为最经典的传统认知诊断模型，其交互函数被广泛用于 DCD 建模中；模糊认知诊断(Fuzzy Cognitive Diagnosis，FCD)框架[8]使用类 IRT 的高阶逻辑函数实现学习者知识掌握水平的模糊化表征，实现对主观题的模糊诊断；深度项目反应理论(Deep Item Response Theory，DIRT)模型[9]使用多种神经网络挖掘题目文本以及其与知识概念之间的关系，增强了 IRT 的诊断过程；神经认知诊断(Neural Cognitive Diagnosis，NCD)框架[10]基于神经网络设计了类 MIRT 交互函数来学习复杂的学习者-题目交互；知识感知认知诊断(Knowledge-Sensed Cognitive Diagnosis，KSCD)[11]框架学习知识概念之间的内在关系，基于 NCD 的类 MIRT 交互函数设计了新的交互函数。这些 DCD 模型分别在题目类型、题目文本和知识概念等层面改进交互函数，提升模型的诊断效果。但这些 DCD 模型对学习者知识掌握水平建模时，默认假设学习者有足够的作答时间来完全发挥知识掌握水平并得到理论上最高的作答准确率[12]。然而，在不同作答场景中学习者并不总有足够的作答时间，学习者可能会牺牲一定作答准确率换取足够的作答速度来完成一场限时课堂测验。

在认知心理学的研究中，作答时间与准确率的关系是重要的研究领域[12-13]，认知水平与速度-准确率权衡(The Speed-Accuracy Tradeoff，SAT)[14]是影响作答时间与准确率关系的重要因素[15-16]。学习者认知水平越高，作答时间越短、作答准确率越高。速度与准确率权衡指学习者可能会牺牲作答速度来换取作答准确率，也可能会牺牲作答准确率来换取作答速度[17]。在认知建模中，作答速度一般用作答时间作为指标，作答速度越快，作答时间越短，反之亦然，即速度与准确率权衡可被视为作答时间与准确率的权衡[18]。

现有 DCD 模型对知识掌握水平的诊断结果是学习者实际发挥的知识掌握水平(以下简称实际知识掌握水平)，而非理论上能达到的最高的知识掌握水平(以下简称理论知识掌握水平)。忽视学习者作答时间与学习者在作答中的速度与准确率权衡会影响诊断结果。例如，学习者在连续作答数道包含相同知识概念的题目时，有个别题目作答时间不足会导致作答准确率显著低于其他。对于异常作答准确率的理解，IRT 可能解释为题目区分度差异，DINA 可能解释为学习者作答失误(Slip)，关注知识概念的 DCD 模型[11,19]可能解释为知识概念之间关系或权重的差异。此外，在认知诊断研究中，题目参数同样是重要研究对象，其中题目区分度不仅会影响诊断精度，还可作为智能组卷中选题策略的重要指标[20-21]。在基于神经网络的类 IRT 交互函数设计上，大部分 DCD 模型采用题目独热向量嵌入的方法表征题目区分度[9-11]，忽视了其他可能的影响因素，比如题目类型的差异对学习者-题目交互有不同影响[8]。

针对上述问题，本文提出基于速度与准确率权衡的深度认知诊断(Speed-Accuracy Tradeoff-Based Deep Cognitive Diagnostic，SAT-CD)模型，将速度与准确率权衡关系融入认知诊断的过程中，实现对学习者实际与理论知识掌握水平的区分诊断。具体来说，首先基于 SAT 设计类 SAT 动态逻辑回归函数来表征学习者实际与理论知识掌握水平，然后引入作答时间属性与题目类型属性表征题目参数，再通过类 MIRT 交互函数模拟学习者-题目的复杂交互，最后预测学习者作答准确率，验证模型效果。在 3 个公开数据集上预测学习者的作答准确率并与同类模型进行对比，并可视化分析学习者的实际与理论知识掌握水平，为诊断结果提供作答时间层面的解释，最后对比了实际与模型所得题目区分度，验证引入题目类型属性的有效性。

2　相关工作

2.1　传统认知诊断

传统教育测量的经典测验理论(Classical Test Theory，CTT)通过整体得分评估学习者，但其参数估计过度依赖学习者样本，且信度估计精度不高。针对这些局限性，IRT 综合考量学习者认知水平与题目参数，项目“Item”指题目，“Item Response”指学习者在题目上的作答。IRT 常使用逻辑回归(Logistic)函数建模，根据不同参数分为单参数(Rasch)模型、双参数模型与三参数模型。单参数模型只考虑知识概念难度；双参数模型考虑知识概念难度与题目区分度；三参数模型考虑知识概念难度、题目区分度与猜测参数。在 DCD 建模研究中，学者们更多地参考双参数 IRT 模型来设计交互函数，

双参数 IRT 函数定义如下：

$$P(x_{ij}=1|\boldsymbol{\theta}_i,\boldsymbol{\beta}_j,\boldsymbol{\eta}_j)=\frac{1}{1+e^{D\eta_j(\theta_i-\beta_j)}} \tag{1}$$

其中，P 为学习者 i 在题目 j 上作答正确的预测值，$\boldsymbol{\theta}_i$ 为学习者 i 的潜在特质，$\boldsymbol{\beta}_j$ 为题目 j 包含知识概念的难度，$\boldsymbol{\eta}_j$ 为题目 j 的区分度。D 为常数，当 D 取 −0.17 时，函数的概率密度与 IRT 另一个基础函数——正态肩型曲线的差异小于 0.01。

2.2 深度认知诊断

DCD 建模中，基于 IRT 设计交互函数是提高模型预测性能和可解释性的有效方法[9-11]。在 DIRT 模型[9]中，通过问题文本得到题目参数，并将其直接输入到双参数 IRT 中预测学习者的作答准确率。在 NCD[10]框架中，基于神经网络设计了能自动学习学习者-题目复杂交互的类 MIRT 交互函数，函数定义如下：

$$x_{ij}=\boldsymbol{Q}_j\circ(\alpha_i-\beta_j)\times\eta_j \tag{2}$$

$$\boldsymbol{y}_{ij}=\sigma(W_3\sigma(W_2\sigma(W_3x_{ij}^{\mathrm{T}}+b_1)+b_2)+b_3) \tag{3}$$

其中，$\boldsymbol{y}_{ij}$ 同作答正确概率 P，为学习者 i 在题目 j 上作答正确的预测值，$\boldsymbol{Q}_j$ 为专家标注的题目-知识概念关联矩阵，$\boldsymbol{\alpha}_i$ 不同于 IRT 中学习者 i 的潜在特质 $\boldsymbol{\theta}_i$，为学习者 i 对知识概念的掌握水平，η_j 和 β_j 分别表示题目 j 区分度与包含的知识概念难度，σ 为 sigmoid 函数，W 和 b 分别为 σ 的权重与偏置参数。

在 NCD 扩展模型文本内容增强的神经认知诊断（Content enhanced NeuralCD with text factor，CNCD-F）模型[10]中，交互函数第一层定义如下（F_j 为题目文本特征）：

$$x_{ij}=\boldsymbol{Q}_j\circ(\alpha_i-(\beta_j\parallel F_j))\times\eta_j \tag{4}$$

在 KSCD 框架[11]中，通过引入知识概念关系，基于式(2)设计了新的交互函数来诊断学习者对非交互知识概念的掌握水平。总而言之，NCD 设计的类 MIRT 交互函数为 DCD 建模提供了一个新的研究方向，即对交互函数进行个性化设计。

2.3 速度与准确率权衡

在实际场景中，学习者与题目的交互比认知建模中的交互更为复杂，并且通常难以用固定学习者在一道题目上的作答时间等控制变量法对速度与准确率权衡进行研究。在真实作答中，学习者往往需要（或希望）在有限的时间内完成作答，此时就需要在作答速度与作答准确率之间进行权衡。学习者的权衡策略会影响在题目上的作答时间，而作答时间是否充足会对作答准确率产生不同程度的影响。

在深度认知诊断建模中，SAT 函数[12,14,22]的定义如下：

$$\boldsymbol{\alpha}_i=\boldsymbol{h}_i\times(1-e^{-\varphi(t_i-\delta)}) \tag{5}$$

其中，$\boldsymbol{\alpha}_i$ 为学习者实际知识掌握水平，$\boldsymbol{t}_i$ 为学习者作答时间，$\boldsymbol{h}_i$ 为学习者在作答时间 $\boldsymbol{t}_i$ 足够长时能达到的理论知识掌握水平，$\boldsymbol{\varphi}$ 为学习者认知加工速率，$\boldsymbol{\delta}$ 为学习者非决策时间（学习者关联题目与知识概念所需时间），$\boldsymbol{t}_i-\boldsymbol{\delta}$ 为学习者有效作答时间。

在 SAT 函数中，实际知识掌握水平 $\boldsymbol{\alpha}_i$ 随着作答时间 $\boldsymbol{t}_i$ 呈指数变化，即学习者作答时间 $\boldsymbol{t}_i$ 越长，实际知识掌握水平 $\boldsymbol{\alpha}_i$ 越接近理论知识掌握水平 $\boldsymbol{h}_i$；$\boldsymbol{h}_i$ 为函数渐近线水平参数，即 $\boldsymbol{\alpha}_i$ 上阈值，$\boldsymbol{\alpha}_i\in(0,\boldsymbol{h}_i]$，表示学习者作答时间 $\boldsymbol{t}_i$ 足够长时实际知识掌握水平 $\boldsymbol{\alpha}_i$ 的最大取值；δ 为函数截距参数，当 $\boldsymbol{\alpha}_i=0$ 时，$\delta=\boldsymbol{t}_i$，即在学习者作答时间至少为 δ 时，才视为开始发挥知识掌握水平；φ 为函数变化速率参数，反映函数曲线陡峭程度，当 $\varphi\to\infty$ 时，$\boldsymbol{\alpha}_i\to\boldsymbol{h}_i$，即学习者认知加工速率 φ 越快，随着作答时间 t_i 增加，实际知识掌握水平 $\boldsymbol{\alpha}_i$ 越快接近理论知识掌握水平 $\boldsymbol{h}_i$（如图 1 所示，$\delta=12$，$\varphi=0.03$，$\boldsymbol{h}_i=1$）。

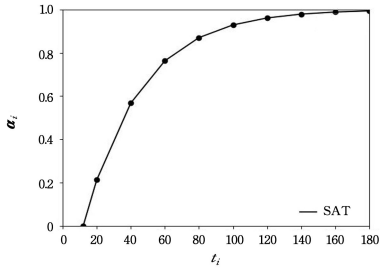

图 1 SAT 函数

Fig. 1 SAT function

在 SAT 研究中的一个基本假设解释了 $\boldsymbol{\alpha}_i$ 下阈值 0 不取闭区间的原因。

假设 1 当学习者作答时间 $\boldsymbol{t}_i>\delta$ 且 $\boldsymbol{t}_i\neq0$ 时，这次作答才被视为一次有效作答。

2.4 认知风格模糊集

不同学习者速度与准确率权衡策略的差异可通过认知风格来解释，认知风格指学习者对外界信息刺激的感知、注意、思维、记忆和解决问题的过程中所偏爱、习惯化的态度和方式[23]。沉思-冲动型二元分类认知风格被广泛接受，该分类解释了学习者作答速度与准确率权衡策略的差异；沉思型学习者倾向于花费更多的时间来作答；而冲动型学习者倾向于快速作答[24]。然而在实际作答场景中，学习者在不同时间作答不同题目时采取的速度与准确率权衡策略可能不同，即学习者认知风格并非恒定不变或二元划分的。

在经典集合中，元素的隶属关系是决定性的二元，诸如生与死、真实与虚假、爱与恨等决定性概念。这些二元概念看似矛盾对立，但在现实中，也有半死不活、半真半假、又爱又恨的模糊概念，因此，仅用二元隶属表征是不够准确的。在深度认知诊断建模中，已有研究通过构建模糊集表征主观题的非对即错的模糊概念[8]，认知风格同样可作为模糊概念进行研究。SAT-CD 引入模糊集（元素具有隶属度的集合）处理学习者认知风格在一定程度上接近沉思型或冲动型等模糊概念。

定义认知风格模糊集为 $(\boldsymbol{S},\boldsymbol{\mu}_i)$，$\boldsymbol{S}$ 为学习者集合，隶属函数为 $\boldsymbol{\mu}_i:\boldsymbol{S}\to(0,1]$（基于假设 1，隶属函数下阈值 0 不取闭区间）。对于每个学习者 $\boldsymbol{s}_i\in\boldsymbol{S}$，定义 $\boldsymbol{s}_i$ 的认知风格 $\boldsymbol{\pi}_i$ 为 $\boldsymbol{s}_i$ 在 $(\boldsymbol{S},\boldsymbol{\mu}_i)$ 中的隶属度 $\boldsymbol{\mu}_i(\boldsymbol{s}_i)$。$\boldsymbol{\mu}_i(\boldsymbol{s}_i)$ 趋近于 0 表示学习者 $\boldsymbol{s}_i$ 的认知风格 $\boldsymbol{\pi}_i$ 倾向于冲动型认知风格，$\boldsymbol{\mu}_i(\boldsymbol{s}_i)$ 趋近于 1 表示学习者 $\boldsymbol{s}_i$ 的认知风格 $\boldsymbol{\pi}_i$ 倾向于沉思型认知风格。

SAT-CD 构建认知风格模糊集旨在通过模糊逻辑将认知风格模糊化为一个值为 $(0,1]$ 的模糊变量，以模拟作答场景中学习者的速度与准确率权衡策略。学习者自身的认知风格会影响学习者作答时采取的速度与准确率权衡策略，并反映在学习者在作答时能将自身理论知识掌握水平实际发挥多少，

即实际知识掌握水平。倾向沉思型认知风格的学习者追求作答时完全发挥自身理论知识掌握水平，即实际知识掌握水平趋近理论知识掌握水平（$\boldsymbol{\alpha}_i \rightarrow \boldsymbol{h}_i$）；而倾向冲动型认知风格的学习者追求快速作答，即实际知识掌握水平不低于知识概念难度（$\boldsymbol{\alpha}_i > \boldsymbol{\beta}_j$）。针对学习者认知风格对速度与准确率权衡策略的影响，本文提出假设 2。

假设 2　学习者认知风格对知识掌握水平的影响在速度与准确率权衡关系中体现为学习者实际知识掌握水平与理论知识掌握水平之比。

3　深度认知诊断建模

3.1　问题描述

在深度认知诊断建模中，定义学习者集合 $\boldsymbol{S}=\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_i\}$、题目集合 $\boldsymbol{E}=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_j\}$、知识概念集合 $\boldsymbol{K}=\{\boldsymbol{k}_1,\boldsymbol{k}_2,\cdots,\boldsymbol{k}_c\}$、题目类型集合 $\boldsymbol{U}=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_d\}$。题目与知识概念的对应关系用题目-知识概念关系矩阵 $\boldsymbol{Q}\in\mathbb{R}^{J\times C}$ 表示，其中 $\boldsymbol{Q}_j\in\{0,1\}^{1\times C}$ 表示题目 $\boldsymbol{e}_j$ 与知识概念 $\boldsymbol{k}_c$ 的对应关系。题目与题目类型的对应关系用题目-题目类型关系矩阵 $\boldsymbol{M}\in\mathbb{R}^{J\times D}$ 表示，其中 $\boldsymbol{m}_j\in\{0,1\}^{1\times D}$ 表示题目 $\boldsymbol{e}_j$ 与题目类型 $\boldsymbol{u}_d$ 的对应关系。学习者在题目上的作答时间用学习者-题目作答时间矩阵 $\boldsymbol{T}\in\mathbb{R}^{I\times J}$ 表示，其中 $\boldsymbol{t}_{ij}$ 表示学习者 $\boldsymbol{s}_i$ 在题目 $\boldsymbol{e}_j$ 上的作答时间。设 $\boldsymbol{R}$ 为学习者题目交互记录，用一组四元组（$\boldsymbol{s}_i,\boldsymbol{e}_j,\boldsymbol{t}_{ij},\boldsymbol{r}_{ij}$）表示，其中 $\boldsymbol{s}_i\in\boldsymbol{S},\boldsymbol{e}_j\in\boldsymbol{E},\boldsymbol{t}_{ij}\in\boldsymbol{T},\boldsymbol{r}_{ij}\in\{0,1\}$ 表示学习者 $\boldsymbol{s}_i$ 在题目 $\boldsymbol{e}_j$ 上的作答结果。

给定学习者集合 $\boldsymbol{S}$、题目集合 $\boldsymbol{E}$、知识概念集合 $\boldsymbol{K}$、题目-知识概念关系矩阵 $\boldsymbol{Q}$、题目-题目类型关系矩阵 $\boldsymbol{M}$、学习者-题目作答时间关系矩阵 $\boldsymbol{T}$ 和题目交互记录 $\boldsymbol{R}$，将学习者知识掌握水平向量 $\boldsymbol{\alpha}_i$ 建模到隐藏空间中，$\boldsymbol{\alpha}_i$ 表示学习者 $\boldsymbol{s}_i$ 对知识概念 $\boldsymbol{k}_c$ 的掌握水平。深度认知诊断建模旨在通过预测学习者题目交互 $\boldsymbol{P}(\boldsymbol{r}_{ij}=1)=\boldsymbol{f}(\boldsymbol{\alpha}_{i,\boldsymbol{\Phi}j})$ 诊断 $\boldsymbol{\alpha}_i$（$\boldsymbol{\Phi}_j$ 为题目参数集，如 IRT 中的知识概念难度 $\boldsymbol{\beta}_k$、题目区分度 $\boldsymbol{\eta}_j$），基于模型预测性能验证所得 $\boldsymbol{\alpha}_i$ 的有效性。SAT-CD 模型结构如图 2 所示，从下往上分别为学习者与题目向量嵌入层、学习者-题目交互层、非负全连接层与输出层。

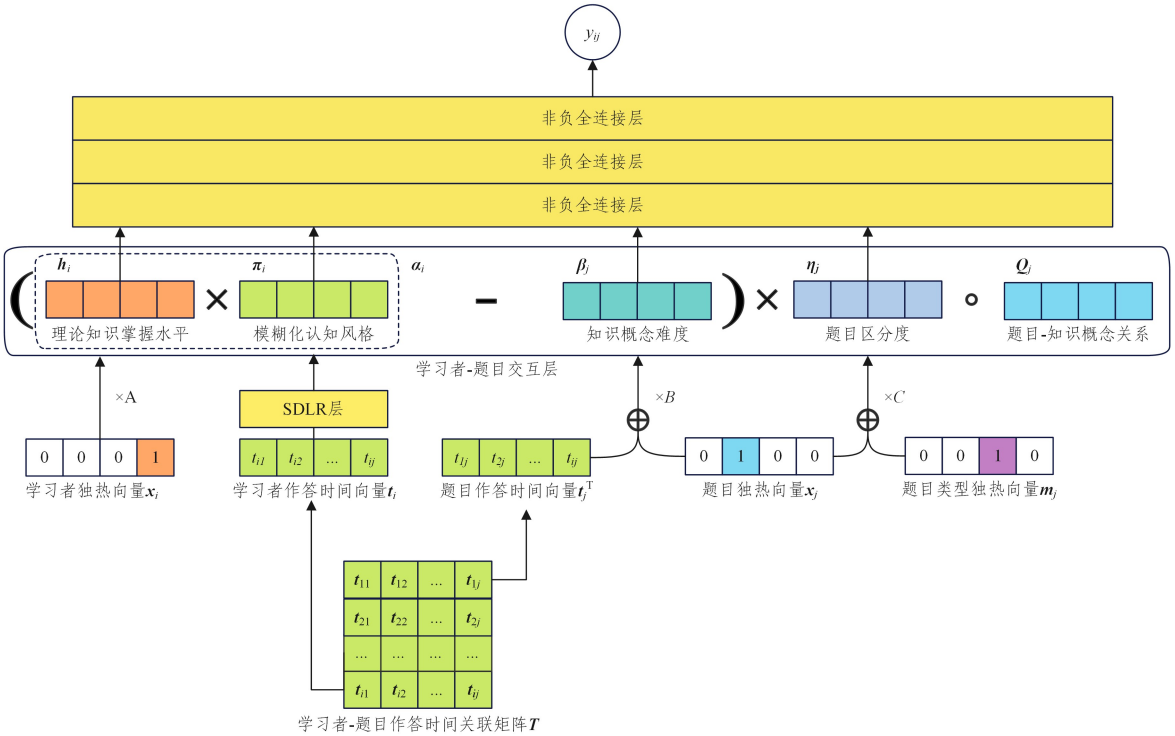

图 2　SAT-CD 模型

Fig. 2　SAT-CD model

3.2　知识掌握水平表征

SAT-CD 基于 SAT 函数的思想设计了类 SAT 动态逻辑回归（SAT-Like Dynamics Logistic Regression，SDLR）函数来表征模糊认知风格 $\boldsymbol{\pi}_i$，函数定义如下：

$$\boldsymbol{\pi}_i=\boldsymbol{\mu}_i(\boldsymbol{s}_i)=\frac{\boldsymbol{\alpha}_i}{\boldsymbol{h}_i}=\frac{1}{1+\boldsymbol{e}^{-\varphi(t_i-\delta)}}+\varepsilon \tag{6}$$

在 SDLR 函数中，学习者的模糊认知风格 $\boldsymbol{\pi}_i$ 随着作答时间 $\boldsymbol{t}_i$ 动态变化，函数形式近似于逻辑回归（Logistic）函数，可训练权重与偏置参数 φ 与 δ 来自于 SAT 函数中的学习者认知加工速率 φ 与学习者非决策时间 δ，学习者作答时间向量 $\boldsymbol{t}_i$ 为学习者-题目作答时间关系矩阵 $\boldsymbol{T}$ 中学习者 $\boldsymbol{s}_i$ 所在行向量。此外，SDLR 函数引入了一个无限接近于 0 但不等于 0 的正数 ε，引入 ε 后 SDLR 函数值域 $(0,1)$ 平移至 $(0+\varepsilon,1+\varepsilon]$，近似于 $(0,1]$，还原 SAT 函数值域的同时也符合假设 1，使模型能更好地模拟学习者的作答速度与准确率权衡策略。

学习者理论知识掌握水平向量 $\boldsymbol{h}_i\in(0,1]\times C$ 由学习者独热向量 $\boldsymbol{x}_i\in\{0,1\}^{1\times I}$ 乘以可训练嵌入矩阵 $\boldsymbol{A}\in\mathbb{R}^{C\times I}$ 嵌入得到：

$$\boldsymbol{h}_i=\boldsymbol{x}_i\times\boldsymbol{A} \tag{7}$$

学习者实际知识掌握水平向量 $\boldsymbol{\alpha}_i\in(0,1]$ 由理论知识掌握水平向量 $\boldsymbol{h}_i$ 与模糊认知风格 $\boldsymbol{\pi}_i$ 相乘得到：

$$\boldsymbol{\alpha}_i = \boldsymbol{\pi}_i \times \boldsymbol{h}_i \tag{8}$$

值得注意的是，如果不经过 SDLR 函数处理，学习者作答时间向量 $\boldsymbol{t}_i$ 作为非独热向量与学习者理论知识掌握水平向量 $\boldsymbol{h}_i$ 相乘或拼接可能会导致维数灾难(Curse of Dimensionality)。具体来说，SDLR 通过将学习者作答时间向量 $\boldsymbol{t}_i$ 映射为模糊认知风格 $\boldsymbol{\pi}_i \in (0,1]$，即使学习者作答时间 $\boldsymbol{t}_i$ 或模糊认知风格 $\boldsymbol{\pi}_i$ 发生动态变化，SDLR 函数也能稳定生成在(0,1]区间内的输出，在有效地控制向量维度的同时避免了维数灾难，增强了 SAT-CD 模型的鲁棒性。

3.3 题目参数表征

3.3.1 知识概念难度表征

在基于 IRT 设计的交互函数中[9-11]，学习者知识掌握水平与对应知识概念难度之差($\boldsymbol{\alpha}_i - \boldsymbol{\beta}_j$)是预测准确率的重要指标，在 3.2 节知识掌握水平表征中，SAT-CD 引入了学习者作答时间 $\boldsymbol{t}_i$ 作为 SDLR 函数的输入属性。类似地，在知识概念难度的表征中，题目作答时间也同样可以作为知识概念难度的特征属性，使交互函数在模拟学习者-题目的复杂交互过程中充分利用学习者作答时间信息。

知识概念难度向量 $\boldsymbol{\beta}_j \in (0,1]^{1\times C}$ 由题目独热向量 $\boldsymbol{x}_j \in \{0,1\}^{1\times J}$ 与学习者-题目作答时间关系矩阵 $\boldsymbol{T}$ 中题目 $\boldsymbol{e}_j$ 所在列向量的转置 $\boldsymbol{t}_j^{\mathrm{T}}$ 拼接嵌入得到：

$$\boldsymbol{\beta}_j = \varepsilon + (1-\varepsilon) \times \sigma((\boldsymbol{x}_j \parallel \boldsymbol{t}_j^{\mathrm{T}}) \times \boldsymbol{B}) \tag{9}$$

其中，$\boldsymbol{B} \in \mathbb{R}^{C\times 2J}$ 为可训练嵌入矩阵、σ 为 sigmoid 函数。为了避免上节提到的维度爆炸问题，知识概念难度向量 $\boldsymbol{\beta}_j$ 表征使用了 sigmoid 激活函数将向量元素映射到(0,1)。由于 SDLR 函数输出的实际知识掌握水平向量 $\boldsymbol{\alpha}_i$ 值域为(0,1]，且在学习者-题目交互函数中，实际知识掌握水平向量 $\boldsymbol{\alpha}_i$ 与知识概念难度向量 $\boldsymbol{\beta}_j$ 进行减法运算，所以在知识概念难度向量 $\boldsymbol{\beta}_j$ 表征中额外引入了 SDLR 函数中的极小正数 ε，使 sigmoid 函数输出值域(0,1)平移至$(0+\varepsilon, 1+\varepsilon]$，近似于(0,1]，且使知识掌握水平与知识概念难度之差($\boldsymbol{\alpha}_i - \boldsymbol{\beta}_j$)保持在(0,1]。

3.3.2 题目区分度表征

题目区分度指题目区分不同知识掌握水平学习者的能力。在基于 IRT 设计的交互函数中[9-11]，题目区分度为预测准确率的权重参数，值域为(0,1)。具体来说，在基于 NCD 框架[10]的衍生模型中，题目区分度标量皆由题目独热向量乘以可训练嵌入矩阵得到，并且知识概念难度也由题目独热向量乘以可训练嵌入矩阵得到。在 KSCD 框架[11]中，题目区分度标量同样由题目独热向量乘以可训练嵌入矩阵得到。这些 DCD 模型没有考量题目类型对题目区分度的影响。在智能组卷与题目设计中，题目类型是影响题目区分度的重要因素，比如主观题与客观题之间、填空题与选择题之间往往具有不同的区分度，不同的题目类型对作答准确率也有较大影响[8,22]。

在 SAT-CD 中，题目区分度的表征引入题目类型属性，题目区分度标量 $\boldsymbol{\eta}_j \in (0,1)$ 由题目独热向量 $\boldsymbol{x}_j \in \{0,1\}^{1\times J}$ 与题目类型独热向量 $\boldsymbol{m}_j \in \{0,1\}^{1\times D}$ 拼接嵌入得到：

$$\boldsymbol{\eta}_j = \sigma((\boldsymbol{x}_j \parallel \boldsymbol{m}_j) \times \boldsymbol{C}) \tag{10}$$

其中，$\boldsymbol{C} \in \mathbb{R}^{1\times(J+D)}$ 为可训练嵌入矩阵、σ 为 sigmoid 函数。引入题目类型属性可以增强 SAT-CD 的泛化能力，模型在遇到新的填空题或选择题时，即使这个题目在训练集中未出现过，模型也能利用已经学习到的关于对应题目类型的知识来判断该题目的区分度。

3.4 学习者-题目交互

3.4.1 交互函数

认知诊断的学习者-题目交互及其函数是认知诊断模型的核心部分，交互函数模拟了学习者与题目交互关系。在传统认知诊断模型[3-4]中，交互函数皆由人工定义，这些专家设计的函数将学习者本身和题目特征线性结合起来，但这不足以捕捉学习者与题目之间更复杂的关系。在 DCD 中，NCD 框架创新性地使用神经网络来建模学习者与题目之间的交互函数，神经网络通过强大的拟合能力逼近任何连续函数，使其能够捕捉学习者与题目属性之间的复杂关系。

通过表征知识掌握水平和题目参数得到学习者实际知识掌握水平向量、知识概念难度向量与题目区分度标量，SAT-CD 使用 NCD[10]设计的基于神经网络的类 MIRT 的交互函数来预测学习者的作答准确率。

NCD 设计的交互函数基于单调性假设。

假设 3 对于任何维度的学习者知识掌握水平，作答准确率都是单调递增的。

这意味着在 DCD 中，假设学习者 $\boldsymbol{s}_i$ 在题目 $\boldsymbol{e}_j$ 上作答正确，在模型训练的过程中，如果模型预测作答错误，模型应该增加(或至少不减)学习者理论知识掌握水平 $\boldsymbol{h}_i$。

学习者-题目交互函数第一层为：

$$\begin{aligned}\boldsymbol{x}_{ij} &= Q_j \circ (\boldsymbol{\alpha}_i - \boldsymbol{\beta}_j) \times \boldsymbol{\eta}_j \\ &= Q_j \circ (\boldsymbol{\pi}_i \circ \boldsymbol{h}_i - \boldsymbol{\beta}_j) \times \boldsymbol{\eta}_j\end{aligned} \tag{11}$$

在交互函数中采用了元素乘$\circ$，该元素乘法可以帮助模型关联学习者实际知识掌握水平与题目-知识概念关系。例如，假设学习者 $\boldsymbol{s}_i$ 在知识概念 $\boldsymbol{k}_c$ 的实际知识掌握水平 $\boldsymbol{\alpha}_i$ 较高，而题目 $\boldsymbol{e}_j$ 包含该知识概念，那么该 $\boldsymbol{s}_i$ 正确回答 $\boldsymbol{e}_j$ 的概率应该会提高。通过元素乘法模型可以将这种关联映射至新向量 $\boldsymbol{x}_{ij}$ 中，进而输入至之后的全连接层中预测学习者作答正确的概率。

预测作答准确率通过两层全连接层与一层输出层实现。

$$\boldsymbol{y}_{ij} = \sigma(W_3\sigma(W_2\sigma(W_1 x_{ij}^{\mathrm{T}} + b_1) + b_2) + b_3) \tag{12}$$

其中，σ 为 sigmoid 函数；W_1, W_2, W_3 为可训练权重参数；b_1，b_2, b_3 为可训练偏置参数。W_1, W_2, W_3 中各元素被设置为非负元素以保证训练过程满足单调性假设[10]。

3.4.2 损失函数

SAT-CD 采用预测作答准确率 $\boldsymbol{y}_{ij}$ 与实际作答准确率 $\boldsymbol{r}_{ij}$ 之间的交叉熵定义损失函数：

$$loss = -\sum_{(s_i, e_j, r_{ij}, r_{ij})} (r_{ij}\log \boldsymbol{y}_{ij}) + (1-\boldsymbol{r}_{ij})\log(1-\boldsymbol{y}_{ij}) \tag{13}$$

4 实验与分析

4.1 预测作答准确率

4.1.1 数据集

通过在 ASSISTments2009－2010(ASSIST2009)，ASSISTments2012－2013(ASSIST2012)，Junyi 这 3 个公开数据集上进行对比实验，验证 SAT-CD 的有效性。ASSIST 数据

集来源于 2004 年创建的 ASSISTments 在线辅导系统[25]。数据集统计信息如表 1 所列。

表 1　数据集统计信息

Table 1　Statistical information of datasets

数据集统计	ASSIST2009	ASSIST2012	Junyi
学习者	4151	25530	116345
题目	16891	42559	713
知识概念	111	241	293
答题记录	346859	440737	1047367

4.1.2　基线模型和评价指标

通过与多个采用类 IRT 交互函数的模型进行对比实验验证 SAT-CD 的有效性，包括基线模型 IRT[3] 与 MIRT[5]、在 IRT 的基础上结合作答时间的统一时间项目反应理论框架(Unified Temporal Item Response Theory，UTIRT)[26]、经典深度认知诊断模型 NCD[10]、在 NCD 基础上加入知识概念关系的 KSCD[11]。使用准确率(ACC)、均方根误差(RMSE)和曲线下面积(AUC)作为评价指标来评估 SAT-CD 在预测学习者作答准确率方面的性能。其中 ACC 是模型对学习者答题的正确预测次数与总预测次数的比值，反映模型的预测能力；RMSE 是观察值(学习者的实际答题结果)与预测值(模型预测的答题结果)之间的均方根误差，反映模型的预测精度；AUC 是 ROC 曲线下的面积，反映模型区分学习者答题正确与否的的能力。

4.1.3　参数设置

SAT-CD 采用 Xavier 初始化来初始化权重参数，根据输入和输出神经元的数量自动调整权重的初始值，从而帮助模型快速收敛。以知识概念数量作为嵌入层的嵌入维度，捕捉学习者知识概念掌握水平，学习者-题目交互函数的全连接层维度分别为 512，256，1，各层均使用 sigmoid 作为激活函数，通过 Adam 优化器训练模型，批量大小(batch size)为 32，学习率(lr)为 0.002。所有数据处理、模型搭建均使用 Pytorch 实现。

4.1.4　对比实验

选取每个模型最好的 5 次预测结果的平均值作为最终结果，如表 2 所列。最好的结果以粗体标示，基线模型中最好的结果以下划线标示(SAT-CD* 相比 SAT-CD 不包括题目类型属性输入)。

表 2　预测实验结果

Table 2　Results of prediction experiment

模型	ASSIST2009			ASSIST2012			Junyi		
	ACC	*RMSE*	*AUC*	*ACC*	*RMSE*	*AUC*	*ACC*	*RMSE*	*AUC*
IRT	0.6744	0.4636	0.6845	0.6634	0.4770	0.6665	0.7133	0.4250	0.7567
MIRT	0.7023	0.4608	0.7189	0.6889	0.4757	0.7043	0.7345	0.4372	0.7657
UTIRT	0.7145	0.4431	0.7372	0.7059	0.4536	0.7317	0.7563	0.4127	0.7982
NCD	0.7205	0.4254	0.7682	0.7154	0.4457	0.7402	0.7533	0.4100	0.7977
KSCD	0.7375	0.4163	0.7722	0.7256	0.4346	0.7478	0.7787	0.3909	0.8090
SAT-CD*	0.7612	0.4046	0.7762	0.7430	0.4123	0.7635	0.8245	0.3432	0.8063
SAT-CD	**0.7633**	**0.4033**	**0.7849**	**0.7542**	**0.4107**	**0.7657**	**0.8305**	**0.3395**	**0.8126**

由实验结果可知：

1)UIRT 相比基线模型 IRT 与 MIRT 预测性能有所提升，证明结合学习者作答时间属性能够帮助模型学习学习者-题目之间的复杂交互。

2)NCD，KSCD 与 SAT-CD* 这 3 个深度认知诊断模型相比模型 IRT，MIRT 与 UTIRT 预测性能皆有提升，证明 NCD 的类 IRT 交互函数能有效学习到学习者-题目之间的复杂交互关系。

3)SAT-CD* 相比 NCD 预测性能显著提升，证明引入速度与准确率权衡关系能够帮助模型学习学习者-题目之间的复杂交互。

4)SAT-CD* 相比 KSCD 预测性能有所提升，在类 IRT 交互函数中，预测作答准确率是基于复杂的学习者-题目交互实现的，证明速度与准确率权衡关系相比知识概念关系能更有效地学习学习者-题目之间的复杂交互。

5)SAT-CD 相比 SAT-CD* 预测性能有所提升，证明引入题目类型属性能帮助模型学习学习者-题目之间的复杂交互。

6)SAT-CD 在 Junyi 数据集上的预测性能提升幅度较大，Junyi 数据集的题目-知识概念关系相比 ASSIST 数据集更稠密，关注知识概念关系的 KSCD 在 Junyi 数据集上难以完全发挥模型优势。

4.2　知识掌握水平分析

不同于现有 DCD 模型对学习者知识掌握水平的单一诊断，SAT-CD 实现了对学习者实际与理论知识掌握水平的区分诊断。以 ASSIST2009 中学习者 s_{71034} 为例，如图 3 所示，学习者在一段时间内作答 8 道包含相同知识概念 k_{85} 的题目 e_j，作答顺序为 e_{85271}，e_{85407}，…，e_{53721}，其中在 e_{84734} 和 e_{53645} 上作答错误。

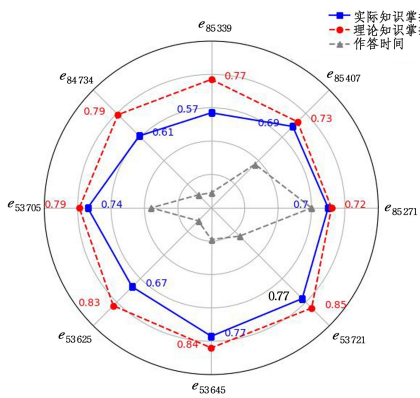

图 3　实际与理论知识掌握水平对比

Fig. 3　Comparison of practical and theoretical knowledge proficiency

学习者理论知识掌握水平 $\boldsymbol{h}_i$ 的诊断满足单调性假设，$\boldsymbol{h}_i \in [0.72, 0.85]$ 随作答序列单调不减。学习者实际知识掌握水平 $\boldsymbol{\alpha}_i$ 的诊断受作答时间 $\boldsymbol{t}_i$ 影响。具体来说，当作答时间较长时，实际知识掌握水平接近于理论知识掌握水平，从而使

作答准确率的期望值增加，比如在较长的作答时间内，s_{71034} 在 e_{85271}，e_{85407}，e_{53705} 上作答正确；当作答时间较短时，实际知识掌握水平与理论知识掌握水平之间存在较大差异，从而使作答准确率的期望值减少，比如在较短的作答时间内，尽管 s_{71034} 在 e_{85339} 和 e_{83734} 上作答正确，但在 e_{53645} 上作答错误。

在诊断结果可解释方面，学习者 s_{71034} 在题目 e_{84734} 上作答错误可解释为作答时间不足，在题目 e_{53645} 上作答错误可解释为题目区分度较高，在题目 e_{85339} 与 e_{53625} 上作答正确可解释为题目区分度较低（在三参数 IRT 或 DINA 中还解释为学习者的失误或猜测因素）。SAT-CD 对实际与理论知识掌握水平的区分诊断有利于帮助学习者自省存在的不足，为预测结果提供作答时间层面的解释。

4.3 题目区分度分析

题目区分度指题目对知识掌握水平不同的学习者的区分程度。假设对一道题目包含的知识概念掌握水平较高的学习者大部分作答正确，掌握水平较低的学习者大部分作答错误，则可认为该题目有较高的区分度。

ASSIST2009 数据集中作答得分为二分变量、作答总分为连续变量，可通过点二列相关法计算题目作答得分与总分的一致性程度来表征该题目的区分度[27]。点二列相关法的基本思想是度量连续变量值如何随二分变量值的改变而改变，具体来说，当二分变量值从一个类变为另一个类时，连续变量的平均值将变化，计算这个变化的大小以及其相对于连续变量的总体标准差的大小可以度量两种变量之间关系的强度。点二列相关函数如下：

$$r_{pb}=\frac{X_p-X_q}{\sigma}\sqrt{\frac{pq}{y}} \tag{14}$$

其中，r_{pb} 为点二列相关系数，表示题目作答得分与总分的一致性程度，p 为该题作答正确的学习者占全体学习者比例，q 为该题作答错误的学习者占全体学习者比例（即 $q=1-p$），X_p 为该题作答正确的学习者总分的平均分，X_q 为该题作答错误的学习者总分的平均分，σ 为总分的标准差。

因为 SAT-CD 诊断所得题目区分度，值域为[0,1]，因此将 r_{pb} 值域[−1,1]映射为[0,1]来表征点二列相关法计算所得题目区分度 η_{pb}：

$$\eta_{pb}=\frac{r_{pb}+1}{2} \tag{15}$$

以 ASSIST2009 中题目 e_{61089}，e_{61091}，e_{61095}，e_{61096}，e_{61097} 为例，通过点二列相关法计算它们的题目区分度，与 SAT-CD 诊断所得题目区分度进行对比，对比结果如表 3 所列。

表 3 实际与诊断题目区分度对比

Table 3 Comparison of actual exercise discrimination and diagnosed exercise discrimination

学习者作答题目	e_{61089}	e_{61091}	e_{61095}	e_{61096}	e_{61097}
作答正确占比 p	0.65	0.8	0.05	0.55	0.85
作答错误占比 q	0.35	0.2	0.95	0.45	0.15
p 总分平均值 X_p	1.62	3.25	1	3.36	3.18
q 总分平均值 X_q	2.14	1.5	2.95	2.44	1.33
题目区分度 η_{pb}	0.35	0.46	0.41	0.49	0.34
题目区分度 η_j	0.31	0.44	0.53	0.23	0.29

题目 e_{61089}，e_{61091}，e_{61097} 的实际区分度与诊断区分度差异较小。在 e_{61089} 中，作答准确率 $p=0.65$，且作答正确的学习者的总分平均值 $X_p=1.62$，低于作答错误学习者的总分平均值 $X_q=2.14$，作答正确的学习者总体知识掌握水平较低，说明 e_{61089} 的区分度较低；在 e_{61091} 与 e_{61097} 中，作答准确率 $p=0.8/0.85$，且作答正确学习者的总分平均值 $X_p=3.25/3.18$，高于作答错误学习者的总分平均值 $X_q=1.5/1.33$，作答正确的学习者总体知识掌握水平较高，而作答错误的学习者总体知识掌握水平较低，说明 e_{61089} 与 e_{61097} 的区分度较高。

题目 e_{61095} 与 e_{61096} 的实际区分度与诊断区分度差异较大。在 e_{61095} 中，作答准确率 $p=0.05$，且作答正确学习者的总分平均值 $X_p=1$，低于作答错误学习者的总分平均值 $X_q=2.95$，作答正确的学习者总体知识掌握水平较低；在 e_{61096} 中，作答准确率 $p=0.55$，且作答正确学习者的总分平均值 $X_p=3.36$，高于作答错误学习者的总分平均值 $X_q=2.44$，作答正确的学习者总体知识掌握水平较高。当题目作答准确率为极值或者中值时，SAT-CD 得到的诊断区分度与实际区分度差异较大，说明 SAT-CD 相比总分平均值更关注作答准确率。

SAT-CD 得到的诊断区分度与点二列相关法计算得到的实际区分度均方根误差为 0.1315，证明了 SAT-CD 引入题目类型表征题目区分度的有效性。在实际应用场景中，比如智能组卷，如果希望学习者在一套题目上的总分符合正态分布，那题目区分度也要符合正态分布[22]。通过对题目区分度的诊断分析，SAT-CD 能有效辅助智能组卷应用制定选题策略。

结束语 本文面向教育测量任务，提出基于速度与准确率权衡的深度认知诊断（SAT-CD）模型，实现对学习者实际与理论知识掌握水平的区分诊断。针对教育测量中限时测验场景，提出类 SAT 动态逻辑回归（SDLR）函数来弥补深度认知诊断在作答时间层面上研究的不足，通过 SDLR 表征实际与理论知识掌握水平，并引入题目类型属性增强模型对复杂交互的学习能力，基于类 MIRT 交互函数构建了深度认知诊断模型。本文进行了学习者的作答准确率预测实验，在公开数据集上验证模型效果，并结合 SAT-CD 的实验结果，从作答时间层面分析实际与理论知识掌握水平，并对比分析了计算所得与诊断所得的题目区分度。

本文的研究重点在于对 SAT 的模拟与实现，在已有的类 MIRT 交互函数上进行扩展，未重新设计交互函数，然而学习学习者-题目的复杂交互关系是认知诊断的重点任务，在类 MIRT 交互函数中，往往将 IRT 中学习者潜在特质表征为知识掌握水平（认知水平）。深度认知诊断模型如何表征学习者作答过程中更高阶的潜在特质（如学习者元认知对自身认知活动的认知调节）值得进一步研究。

参 考 文 献

[1] JIA T, GU X Q. Data technology-driven reshaping of educational forms: paths and processes[J]. China Educational Technology, 2021(3): 38-45.

[2] LUO Z S. Fundamentals of Cognitive Diagnostic Assessment [M]. Beijing: Beijing Normal University Press, 2019.

[3] EMBRETSON S E, REISE S P. Item response theory[M]. Psychology Press, 2013.

[4] DE LA TORRE J. DINA model and parameter estimation：A didactic[J]. Journal of Educational and Behavioral Statistics，2009，34(1)：115-130.

[5] RECKASE M D. The past and future of multidimensional item response theory[J]. Applied Psychological Measurement，1997，21(1)：25-36.

[6] DE LA TORRE J. The generalized DINA model framework[J]. Psychometrika，2011，76：179-199.

[7] LIU Q. Towards a New Generation of Cognitive Diagnosis[C]// IJCAI. 2021：4961-4964.

[8] LIU Q，WU R，CHEN E，et al. Fuzzy cognitive diagnosis for modelling examinee performance[J]. ACM Transactions on Intelligent Systems and Technology(TIST)，2018，9(4)：1-26.

[9] CHENG S，LIU Q，CHEN E，et al. DIRT：Deep learning enhanced item response theory for cognitive diagnosis[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019：2397-2400.

[10] WANG F，LIU Q，CHEN E，et al. NeuralCD：A General Framework for Cognitive Diagnosis[J]. IEEE Transactions on Knowledge and Data Engineering，2023，35(8)：8312-8327.

[11] MA H，LI M，WU L，et al. Knowledge-Sensed Cognitive Diagnosis for Intelligent Education Platforms[C]// Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022：1451-1460.

[12] GUO X J，LUO Z S. Psychometric model and application based on speed-accuracy trade-offs[J]. Psychological Exploration，2019，39(5)：451-460.

[13] WICKELGREN W A. Speed-accuracy tradeoff and information processing dynamics[J]. Acta Psychologica，1977，41(1)：67-85.

[14] HEITZ R P. The speed-accuracy tradeoff：history，physiology，methodology，and behavior[J]. Frontiers in Neuroscience，2014，8：150.

[15] REED A V. Speed-accuracy trade-off in recognition memory[J]. Science，1973，181(4099)：574-576.

[16] DONKIN C，LITTLE D R，HOUPT J W. Assessing the speed-accuracy trade-off effect on the capacity of information processing[J]. Journal of Experimental Psychology：Human Perception and Performance，2014，40(3)：1183.

[17] RATCLIFF R，SMITH P L，BROWN S D，et al. Diffusion decision model：Current issues and history[J]. Trends in Cognitive Sciences，2016，20(4)：260-281.

[18] ZHU Y. Experimental Psychology (2nd ed.)[M]. Beijing：Peking University Press，2009.

[19] ZHANG S J，YU X H，CHEN E H，et al. A Concept Interaction-Based Cognitive Diagnosis Deep Model[J]. Pattern Recognition and Artificial Intelligence，2023，36(1)：22-33.

[20] WANG W Y，SONG L H，DING S L. Cognitive Diagnostic Test Item Distinction Indicators and Applications from a Categorical Perspective[J]. Psychological Science，2018，41(2)：475-483.

[21] HE J，MAO X Z，TANG Q，et al. A dual-objective CD-CAT selection strategy based on item differentiation[J]. Psychological Science，2022，45(1)：204-212.

[22] GUO X J，LUO Z S. The speed-accuracy trade-off：evaluation and modeling of subjects' response states[J]. Studies of Psychology and Behavior，2019，17(5)：589-595.

[23] YAN J H. Cognitive styles affect choice response time and accuracy[J]. Personality and Individual Differences，2010(6)：747-751.

[24] SULISAWATI D N，LUTFIYAH L，MURTINASARI F. Difference of mistakes reflective-impulsive students in mathematical problem solving[J]. International Journal of Trends in Mathematics Education Research，2019(2)：101-105.

[25] FENG M，HEFFERNAN N，KOEDINGER K. Addressing the assessment challenge with an online system that tutors as it assesses[J]. User Modeling and User-Adapted Interaction，2009，19(3)：243-266.

[26] LIU J Y，WANG F，MA H P，et al. A Probabilistic Framework for Temporal Cognitive Diagnosis in Online Learning Systems[J]. Journal of Computer Science and Technology，2023，38(6)：1203-1222.

[27] LI Y H，YANG X Y. Educational and Psychological Statistics[M]. Jiangxi University Press，2020.

CHEN Yan，born in 1976，Ph.D，professor，Ph.D supervisor. Her main reaserch interests include artificial intelligence education，deep learning and emotional computing，artificial intelligence and big data，intelligent information processing，etc.

（责任编辑：何杨）