

# 面向慕课学习者评估的认知反应模型

王炼红<sup>1</sup>, 罗志辉<sup>1</sup>, 刘 畅<sup>2</sup>

(1. 湖南大学电气与信息工程学院, 湖南长沙 410082; 2. 中国农业银行研发中心, 天津 300392)

**摘要:** 认知诊断模型从学习者的认知结构出发, 建模学习者与试题之间的潜在关系, 结合智能算法并根据试题作答结果可评估学习者的知识水平. 大多数认知诊断模型是将学习者的高阶能力特征视为单维, 忽视了后天努力的影响. 为此, 本文提出了一种考虑能力特征与努力特征相互补偿的具有二维高阶特征的新认知诊断模型——认知反应模型(Cognitive and Response Model, C&RM). 该模型通过设置能力特征与努力特征相互补偿机制来融合两高阶特征参数以精准建模学习者的知识水平. 同时, 还构建了知识点弱项特征参数, 以综合考虑学习者的知识水平与不同知识点对作答试题的影响, 进一步提高模型的解释性和预测精度. 采用自建的HNU\_SYS数据集和Math1, Math2, FrcSub公共数据集, 通过实验对比分析了C&RM模型、最新的认知诊断模型和经典诊断模型. 当数据训练集为70%最佳比例时, C&RM在4个数据集上分别比次优方法提升了6.3%, 4.3%, 3.3%, 5.2%, 其预测性能最佳, 验证了本文模型的可行性和有效性.

**关键词:** 认知诊断; 认知反应模型; 评估; 慕课; 补偿机制

**基金项目:** 国家重点研发计划(No.2019YFE0105300); 中国高等教育学会数字化课程资源专项研究课题(No.21SZYB15)

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112(2023)01-0018-08

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20211580

## Cognitive and Response Model for Evaluation of MOOC Learners

WANG Lian-hong<sup>1</sup>, LUO Zhi-hui<sup>1</sup>, LIU Chang<sup>2</sup>

(1. School of Electrical and Information Engineering, Hunan University, Changsha, Hunan 410082, China;

2. R&D Center, Agricultural Bank of China, Tianjin 300392, China)

**Abstract:** The cognitive diagnosis model starts from the learner's cognitive structure, models the potential relationship between the learner and the test questions, and combines intelligent algorithms to evaluate the learner's knowledge level according to the results of the test questions. Most cognitive diagnostic models treat learners' higher-order ability characteristics as a single dimension, ignoring the effect of acquired effort. To this end, this paper proposes a cognitive diagnostic model with two-dimensional high-order features that considers the mutual compensation of ability and effort features—cognitive and response model (C&RM). The model integrates two high-order feature parameters by setting the mutual compensation mechanism of ability feature and effort feature to accurately model the knowledge level of the learner. At the same time, the characteristic parameters of knowledge point weaknesses are also constructed to comprehensively consider the knowledge point level of learners and the influence of different knowledge points on answering questions, and further improve the interpretability and prediction accuracy of the model. Using the self-built HNU\_SYS data set and the Math1, Math2, FrcSub public data sets, the C&RM model, the latest cognitive diagnostic model and the classic diagnostic model are compared and analyzed through experiments. When the data training set is at the best ratio of 70%, C&RM is improved by 6.3%, 4.3%, 3.3%, and 5.2% on the four data sets, respectively, and its prediction performance is the best, which verifies the feasibility of the model in this paper.

**Key words:** cognitive diagnosis; cognitive and response model; evaluation; mooc; compensation mechanism

**Foundation Item(s):** National Key R&D Program of China (No.2019YFE0105300); Special Research Project on Digital Curriculum Resources of China Association of Higher Education (No.21SZYB15)

## 1 引言

学习者习题表现预测旨在根据学习者的历史习题作答记录预测学生未来习题的表现. 近年来,随着教育信息化的迅猛发展,涌现了MOOCs等一系列在线教学平台,在线教学平台可为学习者定制个性化学习方案<sup>[1]</sup>,实现以学习者为中心的个性化学习. 其中,学生的习题表现预测是个性化学习方案定制的基本任务,成为教育数据挖掘领域的研究热点.

认知诊断模型使用数学模型描述学习者知识水平、试题特性与学习者的试题作答结果之间的交互关系. 通过模型求解算法,可以在学习者的试题作答结果以及试题的部分特性已知的条件下,评估学习者的知识水平,根据评估结果预测习题作答表现,进而对学习者的个性化教育资源推荐,节约学习者时间成本<sup>[2-4]</sup>.

现有的认知诊断模型种类繁多,根据模型特点基本可以分为三类:第一类是低阶认知诊断模型,主要有项目反应理论(Item Response Theory, IRT)<sup>[5]</sup>、确定输入噪音与门模型(Deterministic Inputs, Noisy And gate model, DINA)<sup>[6]</sup>、融合模型(Fusion Model, FM)<sup>[7]</sup>、属性层级模型(Attribute Hierarchy Model, AHM)<sup>[8]</sup>等. 该类模型仅对学习者的知识水平与试题作答结果之间的关系进行建模,存在以下问题:(1)将学生的认知过程简单建模为知识决定作答结果的二阶模型,忽视学生和习题本身的特征;(2)模型结构简单,参数空间庞大,容易出现过拟合现象,预测精度不高.

第二类是高阶认知诊断模型<sup>[9-14]</sup>. 为了使模型架构更符合学生的真实认知结构,高阶模型认为学习者不同知识点水平之间是相关的,各知识水平由学习者的高阶能力特征所主导. 高阶 DINA 模型(Higher-Order DINA model, HO-DINA)<sup>[12]</sup>、高阶 IRT 模型<sup>[14]</sup>是其典型代表,都是采用一维高阶能力特征. 高阶 DINA 模型假设学生的知识掌握程度为二元变量,缺乏细粒度的量化过程;高阶 IRT 模型将学生能力到知识水平、知识水平到作答结果的过程均建模为逻辑斯蒂函数,缺乏对个性特征的挖掘. 为了解决高阶 DINA 模型中知识水平存在缺乏细粒度划分的问题, Qi Liu<sup>[13]</sup>基于模糊集理论构建了模糊认知诊断框架(Fuzzy Cognitive Diagnosis Framework, FuzzyCDF),通过模糊化知识水平来建模知识掌握过程.

李忧喜<sup>[9]</sup>认为知识点在题库中的重要性影响了学习者的知识掌握过程,在 FuzzyCDF 基础上,构建了 R-FuzzyCDF. R-FuzzyCDF 将知识点重要性参数定义为知识点出现频次,针对单个试题而言不具备辨别性. 教育心理学研究表明:同一知识点对作答不同试题的重要性也可能是不同的.

第三类是结合神经网络的认知诊断模型<sup>[15,16]</sup>,这类模型使用神经网络来模拟学习者的认知结构,并对神经网络中的训练参数进行约束、赋予实际意义. 通常情况下,这类模型需要大规模的数据进行训练才能取得较好的预测效果,并且由于神经网络的“黑箱”特性,容易出现过拟合.

针对上述模型中的不足,本文提出了一种面向慕课数据等中小规模数据集的高阶认知诊断模型——认知反应模型(Cognitive and Response Model, C&RM). C&RM 的改进创新之处在于:

(1)将一维高阶能力特征扩展为相互独立的能力特征与努力特征,并设置了两者之间的联合补偿机制,基于此机制对学习者的知识掌握过程建模;

(2)考虑学习者的知识点水平与知识点对作答试题影响的重要性差异,引入知识点弱项特征参数建模学习者试题作答过程,进一步提高模型的解释性与预测精度.

## 2 相关工作

### 2.1 低阶认知诊断模型

#### 2.1.1 IRT

IRT 模型表示为

$$P(r_{ij}=1|\theta_i, a_j, d_j, c_j) = c_j + \frac{1 - c_j}{1 + \exp(-Da_j(\theta_i - d_j))} \quad (1)$$

式(1)中,  $r_{ij}$  表示学习者  $i$  对试题  $j$  的作答结果,  $r_{ij}=0$  表示作答错误,  $r_{ij}=1$  表示作答正确.  $P(r_{ij}=1)$  表示学习者  $i$  正确作答试题  $j$  的概率.  $\theta_i$  表示学习者  $i$  的知识水平.  $a_j$  为试题  $j$  的区分度参数,表示试题区分不同知识水平的学生的能力,  $d_j$  为试题  $j$  的难度参数,  $c_j$  为试题  $j$  的猜测度参数,表示学习者通过猜测行为答对试题的概率.  $D$  为尺度常数,通常取 1.7. IRT 模型对学习知识水平的表征是单维的,代表整体的知识水平.

#### 2.1.2 DINA 模型

DINA 模型表示为

$$P(r_{ij}=1|\delta_{ij}, s_j, g_j) = (1 - s_j)^{\delta_{ij}} g_j^{(1 - \delta_{ij})} \quad (2)$$

$$\delta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \quad (3)$$

式(2)中,  $\delta_{ij}$  表示学习者  $i$  是否掌握试题  $j$ .  $s_j, g_j$  分别为试题  $j$  的粗心参数与猜测参数.  $s_j$  表示学习者  $i$  在掌握试题  $j$  的情况下,因粗心导致错误作答的概率,  $g_j$  表示学习者  $i$  在未掌握试题  $j$  的情况下,通过猜测行为正确作答的概率.

式(3)中,  $q_{jk}$  表示试题  $j$  是否考察知识点  $k$ . 并且,  $\delta_{ij}, \alpha_{ik}, q_{jk}$  取值为 0 或 1 的二元类型,即只有掌握和未掌握,或考察和未考察两种状态.

## 2.2 高阶认知诊断模型

### 2.2.1 HO-DINA 模型

HO-DINA 模型如式(4)所示:

$$P(\alpha_{ik}=1)=\frac{1}{1+\exp(-1.7\lambda_{1k}(\theta_i-\lambda_{0k}))} \quad (4)$$

式(4)中,  $\alpha_{ik}$  为式(3)中的知识水平参数.  $P(\alpha_{ik}=1)$  表示学习者  $i$  完全掌握知识点  $k$  的概率,  $\theta_i$  表示学习者  $i$  的高阶能力特征,  $\lambda_{1k}$  表示知识点  $k$  的区分度,  $\lambda_{0k}$  表示知识点  $k$  的难度. 实际上, 学习者的高阶能力可以分为先天能力与后天努力, 但是 HO-DINA 模型中, 仅使用单维值来表示学习者的高阶能力特征.

### 2.2.2 模糊认知诊断框架

FuzzyCDF 在 HO-DINA 模型的基础上, 对式(4)进行了改进, 将学习者的知识水平参数由 0 或 1 的二值变量模糊化为 0-1 之间的模糊变量, 如式(5)所示:

$$\alpha_{ik}=\frac{1}{1+\exp(-1.7\lambda_{1k}(\theta_i-\lambda_{0k}))} \quad (5)$$

式(5)与式(4)中所对应参数含义一致. 此外, FuzzyCDF 使用式(6)来代替式(3)建模知识水平参数  $\alpha_{ik}$  对试题掌握程度  $\delta_{ij}$  的联结型作用<sup>[17]</sup>:

$$\delta_{ij}=\min(\alpha_{i1}, \dots, \alpha_{ik}, \dots, \alpha_{iK}), k \text{ 满足 } q_{jk}=1 \quad (6)$$

FuzzyCDF 认为试题对多知识点是联合考察的, 因此学习者对试题的掌握程度取决于学习者掌握程度最低的知识水平. 实际上, 学习者掌握程度低的知识点可能对该试题作答并不重要, 因此还需要考虑各知识点对作答试题的重要性.

### 2.3 NeuralCD 模型

NeuralCD 模型使用神经网络模拟学习者与试题之间的交互关系. 神经网络的输入为学习者和试题的 one-hot 向量  $\mathbf{x}^s$  和  $\mathbf{x}^e$ , 输出为学习者答对该试题的预测概率  $y$ .

神经网络的输入层表示为

$$\mathbf{x}=\mathbf{Q}_e \circ (\mathbf{h}^s-\mathbf{h}^{\text{diff}}) \times \mathbf{h}^{\text{disc}} \quad (7)$$

$$\mathbf{h}^s=\text{sigmoid}(\mathbf{x}^s \times \mathbf{A}) \quad (8)$$

$$\mathbf{h}^{\text{diff}}=\text{sigmoid}(\mathbf{x}^e \times \mathbf{B}) \quad (9)$$

$$\mathbf{h}^{\text{disc}}=\text{sigmoid}(\mathbf{x}^e \times \mathbf{D}) \quad (10)$$

式(7)中,  $\mathbf{h}^s$  为学习者的知识水平向量, 表示学习者对每个知识点的掌握程度,  $\mathbf{A}$ 、 $\mathbf{B}$ 、 $\mathbf{D}$  均为神经网络的待训练参数矩阵.  $\mathbf{h}^{\text{diff}}$  为试题的知识点难度向量, 表示试题在每个知识点上的考察难度;  $\mathbf{h}^{\text{disc}}$  为试题的区分度参数;  $\mathbf{Q}_e$  为试题与知识点的关联向量.  $\mathbf{x}$  为该层的输出向量. 神经网络的隐层到输出层均为全连接层, 并使用 sigmoid 函数作为激活函数.

与上面提到的认知诊断模型基于统计学理论的求解算法不同, NeuralCD 模型采用梯度下降法求解参数,

求解过程中缺少对现有先验信息的使用, 评估结果缺乏说服力, 且容易出现过拟合.

## 3 本文模型构建

通过对上述模型分析, 我们发现高阶模型较低阶模型结构更为完善, 鲁棒性更好. 本文在高阶模型基础上构建了引入知识点弱项特征参数的具有二维高阶特征的新型认知诊断模型(C&RM).

### 3.1 C&RM 结构

C&RM 模型(见图1)将学习者的认知结构分为四个层次: 二维高阶特征、知识水平、试题掌握程度、试题作答结果. 高阶特征包含学习者的能力与努力特征. 在知识掌握阶段, 学习者的知识水平  $\alpha$  由学习者的二维高阶特征、知识特性参数  $\Psi$  联合确定. 在试题作答阶段, 学习者的知识水平及部分习题特性联合决定了学习者对试题的掌握程度, 再由学习者对试题的掌握程度得到学习者对试题的作答结果.

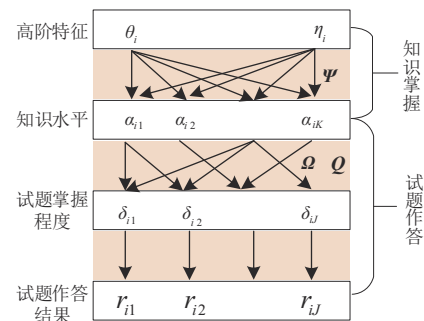


图1 C&RM 结构图

图1中,  $\theta_i$ 、 $\eta_i$  分别表示学习者  $i$  的能力特征、努力特征.  $\Psi$  为知识特性矩阵,  $\Psi=[\Psi_1, \dots, \Psi_K]^T$ ,  $\Psi_k$  表示知识点  $k$  的特征向量, 每个知识点有多个特征参数,  $K$  表示课程知识点数量.  $\alpha_{ik}$  表示学习者  $i$  对知识点  $k$  的知识水平.  $\Omega$  为试题特性矩阵,  $\Omega=[\Omega_1, \dots, \Omega_J]^T$ ,  $\Omega_j$  表示试题  $j$  的特征向量, 每道试题有多个特征参数,  $J$  为试题的数量.  $\mathbf{Q}$  为试题-知识点关联矩阵, 由一系列试题-知识点关联向量构成,  $\mathbf{Q}=[\mathbf{q}_1, \dots, \mathbf{q}_J]^T$ ,  $\mathbf{q}_j=[q_{j1}, \dots, q_{jK}]$ ,  $q_{jk}$  取值为二元类型,  $q_{jk}=1$  表示试题  $j$  考察了知识点  $k$ ,  $q_{jk}=0$  表示试题  $j$  未考察知识点  $k$ .  $\delta_{ij}$  表示学习者  $i$  对试题  $j$  的掌握程度.  $r_{ij}$  表示学习者  $i$  对试题  $j$  的作答结果, 取值为二元类型,  $r_{ij}=1$  表示作答正确,  $r_{ij}=0$  表示作答错误.

### 3.2 能力与努力特征相互补偿机制及知识掌握过程建模

#### 3.2.1 能力与努力特征互相补偿机制

对学习者的能力与努力特征之间的联合补偿机制设置如下:

(1) 能力特征与努力特征对学习者的知识水平的影

响效果是可以相互补偿的. 能力低的学习者可以通过努力达到和能力高的学习者同样的知识水平.

(2) 存在于能力特征与努力特征之间的补偿效果是有限的. 某些能力或努力程度低于一定范围的学习者, 即使可以通过补偿取得补偿效果, 但该种补偿效果具备上限.

(3) 学习者的知识水平由能力特征与努力特征联合决定. 在极端情况下, 当能力特征或努力特征趋近于最小值时, 学习者的知识水平也应当趋近于最小值.

### 3.2.2 学习者知识掌握过程建模

考虑上述补偿机制, 对学习者的知识掌握过程建模如下:

$$\alpha_{ik} = \frac{\beta_{\theta k} w_{k1}}{1 + \exp(-D \cdot (\theta_i - d_{\theta k1}))} + \frac{\beta_{\eta k} w_{k2}}{1 + \exp(-D \cdot (\eta_i - d_{\eta k1}))} \quad (11)$$

$$\beta_{\theta k} = \frac{1}{1 + w_{k2} \exp(-D \cdot (\eta_i - d_{\eta k2}))} \quad (12)$$

$$\beta_{\eta k} = \frac{1}{1 + w_{k1} \exp(-D \cdot (\theta_i - d_{\theta k2}))} \quad (13)$$

$$w_{k1} + w_{k2} = 1 \quad (14)$$

式(11)~(14)中,  $w_{k1}$ 、 $w_{k2}$  分别表示知识点  $k$  对学习者的能力特征、努力特征的考察权重. 当  $w_{k1}$  或  $w_{k2}$  为 0 时, 模型退化为只考虑学习者单维高阶特征.  $\beta_{\theta k}$ 、 $\beta_{\eta k}$  分别为能力特征及努力特征对知识水平作用效果的衰减系数. 以  $\beta_{\theta k}$  为例说明, 它是能力特征对知识水平作用效果的衰减系数, 由式(12)可看出,  $\beta_{\theta k}$  与努力特征权重  $w_{k2}$  负相关, 与努力特征  $\eta_i$  正相关, 这是因为知识的掌握是对能力、努力两种特征联合考察的, 当努力较低的时候, 能力特征对知识水平的原本作用效果会出现衰减, 并且努力特征越重要, 衰减效果越严重.  $d_{\theta k1}$ 、 $d_{\theta k2}$ 、 $d_{\eta k1}$ 、 $d_{\eta k2}$  表示知识点  $k$  在能力特征及努力特征上的难度参数.  $D$  为常数, 取值 1.7.

### 3.3 引入知识点弱项参数的试题作答过程建模

#### 3.3.1 学习者试题掌握程度

C&RM 中, 学习者对试题的掌握程度  $\delta_{ij}$  依赖于学习者对试题考察的知识点的弱项参数  $l_{ijk}$ :

$$\delta_{ij} = 1 - \max(l_{ij1}, \dots, l_{ijk}), \quad k \text{ 满足 } q_{jk} = 1 \quad (15)$$

$$l_{ijk} = e_{jk} (1 - \alpha_{ik}) \quad (16)$$

式(16)中, 知识点弱项参数  $l_{ijk}$  为知识点  $k$  对作答试题  $j$  的重要性  $e_{jk}$  与学习者  $i$  对知识点  $k$  未掌握程度的乘积. 即知识点  $k$  在作答试题  $j$  时越重要且学习者  $i$  对知识点  $k$  的掌握程度越低, 那么知识点  $k$  构成知识弱项的程度就越高.

#### 3.3.2 学习者试题作答结果

C&RM 中, 学习者对试题的掌握程度与学习者的

试题作答结果关系如下:

$$P(r_{ij} = 1) = \frac{1}{1 + \exp(-1.7\lambda_{ij}(\delta_{ij} - \lambda_{0j}))} \quad (17)$$

式(17)中,  $\delta_{ij}$  表示学习者  $i$  对试题  $j$  的掌握程度,  $\lambda_{ij}$ 、 $\lambda_{0j}$  分别表示试题  $j$  对学习者的试题掌握程度的区分度以及难度.

### 3.4 基于 MCMC 的 C&RM 参数估计

采用 C&RM 对学习者的评估时, 学习者的高阶努力特征、试题-知识点关联矩阵以及试题作答结果是已知参数, 学习者的高阶能力特征、知识水平, 知识特征向量与试题特征向量是待求解参数. 其中, 根据 3.2 节、3.3 节可知: 知识特征向量为  $\Psi_k = [w_{k1}, w_{k2}, d_{\theta k1}, d_{\eta k1}, d_{\theta k2}, d_{\eta k2}]$ ; 试题特征向量为  $\Omega_j = [\lambda_{0j}, \lambda_{1j}, e_j]$ . 由式(11)~(14)各参数之间的关系可知, 只需对参数  $\theta$ 、 $\Psi$ 、 $\Omega$  进行求解, 其他参数  $\alpha$ 、 $\delta$  就可相继求出.

本文采用马尔可夫蒙泰卡罗算法 (Markov Chain Monte Carlo, MCMC) [18] 对参数  $\theta$ 、 $\Psi$ 、 $\Omega$  进行求解. MCMC 算法能简化参数估计的复杂度, 可满足多维模型、多类参数的估计任务, 且估计精度高, 稳定性好.

根据贝叶斯公式及式(11)~(14), 计算参数  $\Psi$ 、 $\theta$ 、 $\Omega$  的满条件后验分布分别为

$$P(\Psi|\theta, \Omega, r) \propto \prod_{k=1}^K P(w_{k1})P(d_{\theta k1})P(d_{\eta k1}) \times P(d_{\theta k2}|d_{\theta k1})P(d_{\eta k2}|d_{\eta k1}) \times \prod_{i=1}^I \prod_{j=1}^J (P(r_{ij} = 1|\theta, \Omega, \Psi))^{r_{ij}} (1 - P(r_{ij} = 1|\theta, \Omega, \Psi))^{1-r_{ij}} \quad (18)$$

$$P(\theta|\Psi, \Omega, r) \propto \prod_{i=1}^I P(\theta_i) \times \prod_{i=1}^I \prod_{j=1}^J (P(r_{ij} = 1|\theta, \Omega, \Psi))^{r_{ij}} (1 - P(r_{ij} = 1|\theta, \Omega, \Psi))^{1-r_{ij}} \quad (19)$$

$$P(\Omega|\Psi, \theta, r) \propto \prod_{j=1}^J P(e_j)P(\lambda_{0j})P(\lambda_{1j}) \times \prod_{i=1}^I \prod_{j=1}^J (P(r_{ij} = 1|\theta, \Omega, \Psi))^{r_{ij}} (1 - P(r_{ij} = 1|\theta, \Omega, \Psi))^{1-r_{ij}} \quad (20)$$

参数求解算法见算法 1.

参数  $\Psi^* = [\Psi_1^*, \dots, \Psi_K^*]^T$ ,  $\theta^* = [\theta_1^*, \dots, \theta_I^*]$ ,  $\Omega^* = [\Omega_1^*, \dots, \Omega_J^*]$  中各元素具体抽样过程如下:

#### (1) $\Psi_k^*$ 参数

从均匀分布  $U(w_{k1}' - \delta_w, w_{k1}' + \delta_w)$ , 正态分布  $N(d_{\theta k1}' - 1, \sigma_{d_{\theta k1}}^2)$ ,  $N(d_{\eta k1}' - 1, \sigma_{d_{\eta k1}}^2)$  中随机抽样得到  $w_{k1}^*$ ,  $d_{\theta k1}^*$ ,  $d_{\eta k1}^*$ , 然后根据式(11)及 4.2 节实验参数设置计算得到  $w_{k2}^*$ ,  $d_{\theta k2}^*$ ,  $d_{\eta k2}^*$ .

**算法 1** 马尔可夫蒙泰卡罗算法输入:  $\mathcal{Q}$ 、 $r$ 、 $\eta$ 输出:  $\Psi$ 、 $\theta$ 、 $\Omega$ 随机初始化参数初始值  $\Psi^1$ 、 $\theta^1$ 、 $\Omega^1$ FOR  $t=1, \dots, T$     抽样得到  $\Psi^*$ , 根据参数  $\Psi$  的满条件分布计算  $P(\Psi^*, \Psi^*) =$      $\min\left\{\frac{P(\Psi^*|\theta^t, \Omega^t, r)}{P(\Psi^t|\theta^t, \Omega^t, r)}, 1\right\}$ , 并以  $P(\Psi^*, \Psi^*)$  的概率接受  $\Psi^*$  为  $\Psi^{t+1}$ , 否则     $\Psi^{t+1} = \Psi^t$ .    抽样得到  $\theta^*$ , 根据参数  $\theta$  的满条件分布计算  $P(\theta^*, \theta^*) =$      $\min\left\{\frac{P(\theta^*|\Omega^t, \Psi^{t+1}, r)}{P(\theta^t|\Omega^t, \Psi^{t+1}, r)}, 1\right\}$ , 并以  $P(\theta^*, \theta^*)$  的概率接受  $\theta^*$  为  $\theta^{t+1}$ , 否则     $\theta^{t+1} = \theta^t$ .    抽样得到  $\Omega^*$ , 根据参数  $\Omega$  的满条件分布计算  $P(\Omega^*, \Omega^*) =$      $\min\left\{\frac{P(\Omega^*|\theta^{t+1}, \Psi^{t+1}, r)}{P(\Omega^t|\theta^{t+1}, \Psi^{t+1}, r)}, 1\right\}$ , 并以  $P(\Omega^*, \Omega^*)$  的概率接受  $\Omega^*$  为  $\Omega^{t+1}$ , 否    则  $\Omega^{t+1} = \Omega^t$ .

END

(2)  $\theta_i^*$  参数从正态分布  $N(\theta_i^t, \sigma_\theta^2)$  中随机抽样得到  $\theta_i^*$ .(3)  $\Omega_j^*$  参数从均匀分布  $U(e_j^t - \delta_e, e_j^t + \delta_e)$ 、 $U(\lambda_{0j}^t - \delta_{\lambda_0}, \lambda_{0j}^t + \delta_{\lambda_0})$ 、 $U(\lambda_{1j}^t - \delta_{\lambda_1}, \lambda_{1j}^t + \delta_{\lambda_1})$  中随机抽样得到  $e_j^*$ 、 $\lambda_{0j}^*$ 、 $\lambda_{1j}^*$ .其中上标  $t$  表示对应参数的第  $t$  次迭代结果 (马尔可夫链执行到  $t$  步时, 该参数的状态值), 参数  $T$ 、 $\delta_w$ 、 $\sigma_{d_e}^2$ 、 $\sigma_{d_\theta}^2$ 、 $\sigma_{d_\lambda}^2$ 、 $\delta_{\lambda_0}$ 、 $\delta_{\lambda_1}$ 、 $\delta_e$  按照 4.2 节实验参数设置取值.算法结束后, 判断参数对应的马尔可夫链的收敛性. 若收敛, 则对该马尔可夫链的平稳期进行采样, 计算采样序列的平均值作为该参数的求解结果; 若不收敛, 则增大  $T$  的取值并继续运行算法直至马尔可夫链收敛为止, 然后再进行采样计算平均值.

## 4 实验

### 4.1 数据集

本文选择 FrcSub<sup>[13]</sup>, Math1<sup>[13]</sup>, Math2<sup>[13]</sup>, HNU\_SYS 四个数据集来进行对比试验. 其中 FrcSub、Math1、Math2 是目前认知诊断模型研究中常用的公开数据集, 数据集包含学习者的试题作答结果以及试题-知识点关联矩阵.

HNU\_SYS 数据集于中国大学 MOOC 课程平台爬虫获取, 并剔除了作答记录少于 10 的用户数据. 该数据集包含了学习者的试题作答结果与试题-知识点关联矩阵信息, 还包含对学习者的努力特征参数的评估结果. 数据集详细信息见表 1. HNU\_SYS 数据集中, 学习者的努力特征参数是对学习者的作业、测验总成绩通过式 (21) 变换得到. 因为作业和测验难度较低, 可以较好的

体现学习者的努力特征.

$$\int_{-\infty}^{\eta_i} f(x) dx = \frac{\text{rank}(i)}{\text{Total} + 1} \quad (21)$$

测试、作业总成绩与努力特征参数变换的原则是保证变换前后学习者  $i$  所在的位次比例不变, 并且变换后其参数服从标准正态分布. 式 (21) 中,  $f(x)$  为标准正态分布的概率密度函数,  $\eta_i$  表示学习者  $i$  的努力特征参数,  $\text{rank}(i)$  为学习者  $i$  作业、测验总成绩在所有学习者中的逆序排名位次. Total 为学习者总人数. 分母 Total + 1 避免了等式右边取值为 1 导致等式左边  $\eta_i$  取值为无穷大的问题.

由于本文提出的 C&RM 对学习者的知识掌握过程与试题作答过程都进行了改进, 为了充分验证两部分改进的有效性, 将 C&RM 分为仅对试题作答过程改进的 C&RM-1 和对知识掌握、试题作答过程都进行改进的 C&RM-2. 首先在 FrcSub, Math1, Math2 数据集上验证 C&RM-1, 然后在 HNU\_SYS 数据集上验证 C&RM-1 与 C&RM-2.

表 1 数据集概要

数据集	学生数	知识点数	试题数
FrcSub	536	8	20
Math1	4 209	11	15
Math2	3 911	16	16
HNU_SYS	466	6	90

### 4.2 实验参数设置

3.4 节参数估计算法中如下参数的先验分布依照文献 [6, 12] 中的经验来确定.

$$\theta_i \sim N(0, 1), w_{k1} \sim U(0, 1), d_{\eta k1} \sim N(0, 1)$$

$$d_{\theta k1} \sim N(0, 1), \lambda_{0j} \sim N(0.5, 0.15), \lambda_{1j} \sim \log N(1, 1), e_{jk} \sim \text{beta}(2, 1)$$

以下参数取值参照文献 [12] 中的经验值设定:

$$\delta_w = 0.01, \sigma_{d_e}^2 = 0.3, \sigma_{d_\theta}^2 = 0.3, \sigma_\theta^2 = 1.1$$

$$\delta_e = 0.05, \delta_{\lambda_1} = 0.05, \delta_{\lambda_0} = 0.05, T = 6000$$

为了简化算法迭代过程, 参数  $d_{\eta k2}$  和  $d_{\theta k2}$  按照如下公式确定.

$$d_{\eta k2} = d_{\eta k1} + \frac{\log(0.001) - \log(w_{k2}) - \log(0.999)}{D}$$

$$d_{\theta k2} = d_{\theta k1} + \frac{\log(0.001) - \log(w_{k1}) - \log(0.999)}{D}$$

参数  $d_{\eta k2}$  表征知识点  $k$  对学习者的努力考察难度, 与知识点  $k$  对学习者的努力特征的考察权重  $w_{k2}$  成反比. 参数  $d_{\theta k2}$  的设置同理, 确保自设参数取值的合理性. 由于参数  $d_{\eta k2}$  和  $d_{\theta k2}$  的取值是自设确定, 为了验证参数对 C&RM 预测效果的影响, 本文于 4.4 小节对上述两参数进行了灵敏度分析.

### 4.3 C&RM 预测对比实验

本文将学习者试题作答记录分为训练集与测试集,基于训练集中的数据评估学习者的知识水平与知识特性、试题特性,基于评估结果对测试集中的试题进行预测.使用正确率(Accuracy)作为评价指标,如式(22)所示.选择 IRT<sup>[5]</sup>,DINA<sup>[6]</sup>,HO-DINA<sup>[12]</sup>,FuzzyCDF<sup>[13]</sup>,R-FuzzyCDF<sup>[9]</sup>,HO-PINC<sup>[10]</sup>,NeuralCD<sup>[15]</sup>作为C&RM的对比模型.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (22)$$

TP表示样本真实标签为正确,预测为正确的习题数;TN表示样本真实标签为错误,预测为错误的习题数;FP表示样本真实标签为错误,预测为正确的习题数;FN表示样本真实标签为正确,预测为错误的习题

数. Accuracy 值越高,表明模型的预测性能越好. Accuracy 值越低,表明模型的预测性能越差.对比实验结果如图2(a)~(d)所示.

图2(a)~(d)中,横坐标为训练集样本占数据集样本的比例,纵坐标为预测正确率指标.由于Frcsub, Math1, Math2数据集无法计算努力特征参数,在上述数据集上C&RM模型采用C&RM-1.

从图2(a)~(c)中可以看出,在FrcSub, Math1, Math2数据集上,C&RM-1预测性能均优于其他模型.从图2(d)可以看出,在自制数据集HNU\_SYS上,C&RM-1和C&RM-2的预测效果均优于其他对比模型,且C&RM-2更优.其中,在两极端的训练比例下(70%、30%),C&RM-1/2对比IRT, DINA, HO-DINA, FuzzyCDF, R-FuzzyCDF, HO-PINC, NeuralCD等算法的提升见表2.

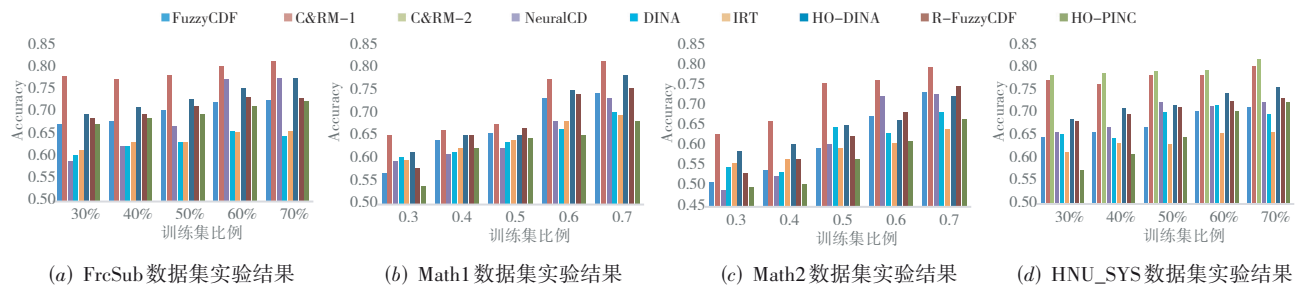


图2 C&RM 预测对比实验结果

表2 HNU\_SYS上C&RM-1/2相对其他模型的提升效果

模型	C&RM-1		C&RM-2	
	70% 训练	30% 训练	70% 训练	30% 训练
IRT	14.7%	15.9%	16.6%	17.1%
DINA	10.7%	10.9%	12.6%	13.1%
HO-DINA	4.7%	8.6%	6.6%	9.8%
FuzzyCDF	9.0%	12.6%	10.9%	13.8%
R-FuzzyCDF	7.2%	9.1%	9.1%	10.3%
HO-PINC	7.9%	19.9%	9.8%	21.1%
NeuralCD	7.8%	11.6%	9.6%	12.8%
C&RM-1	0	0	8.8%	1.2%

实验结果充分验证了C&RM中的改进是切实有效的,说明C&RM的模型结构更接近学习者真实的认知结构,基于C&RM对学习者的评估时说服力更强.

### 4.4 参数灵敏度分析

本文所提出的C&RM中,参数 $d_{\eta k2}$ 和 $d_{\theta k2}$ 作为能力-努力特征融合机制中的关键参数,表征知识点 $k$ 在能力特征及努力特征上的考察难度.且 $d_{\eta k2}$ 的值越大,表示知识点 $k$ 对于学生努力特征的考察难度更大; $d_{\theta k2}$ 的值越大,表示知识点 $k$ 对于学生能力特征的考察难度更大.

由于FrcSub和HNU\_SYS两数据集规模相当,前者

是常用经典标准数据集,后者是自制数据集,两数据集具有代表性,因此选用Frcsub数据集和HNU\_SYS数据集,在70%训练比例下,检验不同的参数取值对于C&RM模型的预测效果影响,实验结果见图3和图4.

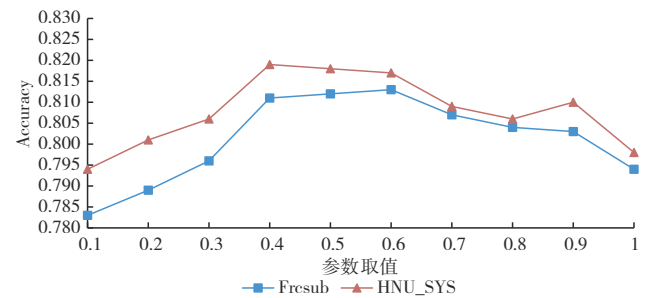


图3  $d_{\eta k2}$ 取值对C&RM预测效果的影响

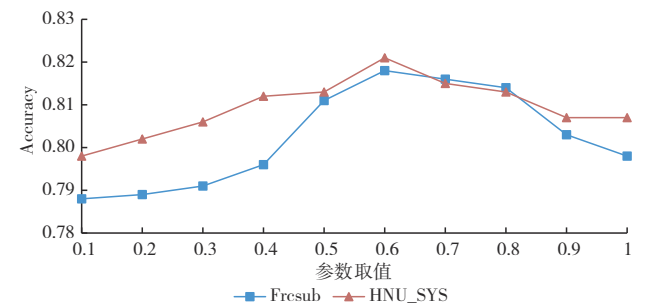


图4  $d_{\theta k2}$ 取值对C&RM预测效果的影响

从图中可以看出,  $d_{\eta k_2}$  和  $d_{\theta k_2}$  的取值对于本文所提模型的预测性能有显著的影响. 对参数  $d_{\eta k_2}$  而言, 当取值为 0.4 时, 在 HNU\_SYS 和 Frsub 上的表现均为最优; 对参数  $d_{\theta k_2}$  而言, 当取值为 0.6 时, 在数据集 HNU\_SYS 和 Frsub 上的表现均为最优. 从结果中不难看出, 两个数据集包含的习题对于学习者努力特征和能力特征的考察程度具备差异性. 此外, 当  $d_{\eta k_2}$  和  $d_{\theta k_2}$  取两端的极端值时, 预测性能显著下降, 从侧面验证了 C&RM 引入努力参数的重要性.

## 5 总结与展望

针对现有认知诊断模型的不足, 本文提出了一种改进的高阶认知诊断模型——C&RM. C&RM 首次提出融合能力特征与努力特征来建模学习者的知识掌握过程, 并结合学习者的知识水平与知识点对作答试题的重要性, 构建了知识点弱项这一特征参数来建模学习者的试题作答结果. C&RM 对学习者的知识掌握、试题作答的过程考虑更加完备, 并且基于评估结果对学习者的试题作答进行预测时, 精度更高, 充分说明了 C&RM 的优越性.

C&RM 应用于慕课教学时, 一方面可以对参与课程的学习者的知识水平进行诊断评估, 帮助学习者查漏补缺; 另一方面还可通过预测学习者对试题的正确作答概率, 从而推荐合理难度的试题供学习者训练提高. 此外, 本文还存在一些不足: 认知诊断算法的初始参数设置有待优化; 学生努力特征参数的可信度有待提高, 这是下阶段工作的重点.

### 参考文献

- [1] 祝智庭, 魏非. 教育信息化 2.0: 智能教育启程, 智慧教育领航[J]. 电化教育研究, 2018, 39(9): 5-16.  
ZHU Zhi-ting, WEI Fei. Educational informatization 2.0: Starting on a journey of intelligence education guided by smart education[J]. e-Education Research, 2018, 39(9): 5-16. (in Chinese)
- [2] 朱天宇, 黄振亚, 陈恩红, 等. 基于认知诊断的个性化试题推荐方法[J]. 计算机学报, 2017, 40(1): 176-191.  
ZHU Tian-yu, HUANG Zhen-ya, CHEN En-hong, et al. Cognitive diagnosis based personalized question recommendation[J]. Chinese Journal of Computers, 2017, 40(1): 176-191. (in Chinese)
- [3] 王超, 刘淇, 陈恩红, 等. 面向大规模认知诊断的 DINA 模型快速计算方法研究[J]. 电子学报, 2018, 46(5): 1047-1055.  
WANG Chao, LIU Qi, CHEN En-hong, et al. The rapid calculation method of dina model for large scale cognitive diagnosis[J]. Acta Electronica Sinica, 2018, 46(5): 1047-1055. (in Chinese)
- [4] LIU H, ZHAO W, LIANG M. Pedagogical strategy model in adaptive learning system focusing on learning styles[C]// International Conference on Technologies for E-Learning and Digital Entertainment. Berlin: Springer, 2010: 156-164.
- [5] FAN X. Item response theory and classical test theory: An empirical comparison of their item/person statistics[J]. Educational & Psychological Measurement, 1998, 58(3): 357-381.
- [6] DE LA TORRE J. DINA model and parameter estimation: A didactic[J]. Journal of Educational and Behavioral Statistics, 2009, 34(1): 115-130.
- [7] Hartz M C. A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality[J]. American Journal of Gastroenterology, 2002, 95(4): 906-909.
- [8] LEIGHTON J P, GIERL M J, HUNKA S M. The attribute hierarchy method for cognitive assessment: A variation on Tatsuoaka's rule-space approach[J]. Journal of Educational Measurement, 2004, 41(3): 205-237.
- [9] 李忧喜, 文益民, 易新河, 等. 一种改进的模糊认知诊断模型[J]. 数据采集与处理, 2017, 32(5): 958-969.  
LI You-xi, WEN Yi-min, YI Xin-he, et al. Revised model of fuzzy cognitive diagnosis framework[J]. Journal of Data Acquisition and Processing, 2017, 32(5): 958-969. (in Chinese)
- [10] ZHAN P, WANG W C, JIAO H, et al. Probabilistic-input, noisy conjunctive models for cognitive diagnosis[J]. Frontiers in Psychology, 2018, 9: 997-1002.
- [11] ZHAN P, MA W, JIAO H, et al. A sequential higher order latent structural model for hierarchical attributes in cognitive diagnostic assessments[J]. Applied Psychological Measurement, 2020, 44(1): 65-83.
- [12] TORRE J, DOUGLAS J A. Higher-order latent trait models for cognitive diagnosis[J]. Psychometrika, 2004, 69(3): 333-353.
- [13] LIU Q, WU R Z, CHEN E H, et al. Fuzzy cognitive diagnosis for modelling examinee performance[J]. ACM Transactions on Intelligent Systems and Technology, 2018, 9(4): 1-26.
- [14] DE L, SONG H. Simultaneous estimation of overall and domain abilities: A higher-order IRT model approach[J]. Applied Psychological Measurement, 2009, 33(8): 620-639.
- [15] WANG F, LIU Q, CHEN E, et al. Neural cognitive diag-

nosis for intelligent education systems[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2020: 6153-6161.

- [16] CHEN J, LIN J, LI X G. Research on cognitive diagnostic model based on BP neural network[C]//2019 4th International Conference on Measurement, Information and Control (ICMIC). Piscataway: IEEE, 2019: 61-65.
- [17] PARDOS Z A, HEFFERNAN N T, RUIZ C, et al. The composition effect: Conjunctive or compensatory? An analysis of multi-skill math questions in ITS[C]//International Conference Educational Data Mining. Las Vegas: IEDMS, 2008: 147-156.
- [18] GELFAND A E, HILLS S E, SMITH R P F M. Illustration of Bayesian inference in normal data models using Gibbs sampling[J]. Journal of the American Statistical Association, 1990, 85(412): 972-985.

#### 作者简介



王炼红 女,1971年5月生,湖南宁乡人. 博士,副教授、硕士生导师. 2011年3月至2012年3月,于美国布兰迪斯大学做访问学者. 主要研究方向为信号处理、数据挖掘技术、人工智能.  
E-mail: wanglh@hnu.edu.cn



罗志辉 男,1998年9月出生,出生于湖南省永州市. 2020年于南京农业大学获得工学学士学位. 现为湖南大学电气与信息工程学院硕士研究生. 主要研究方向为教育数据挖掘和机器学习.  
E-mail: luo1998@hnu.edu.cn