

## 基于注意力机制的概念增强认知诊断模型

苑冬雪, 孙权森, 傅鹏

引用本文

苑冬雪, 孙权森, 傅鹏. 基于注意力机制的概念增强认知诊断模型[J]. 计算机科学, 2023, 50(11): 241-247.

YUAN Dongxue, SUN Quansen, FU Peng. Attention Based Concept Enhanced Cognitive Diagnosis[J]. Computer Science, 2023, 50(11): 241-247.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于贝叶斯规则的具有层次注意力的知识补全](#)

Bayesian Rule-based Knowledge Completion with Hierarchical Attention

计算机科学, 2023, 50(11): 234-240. <https://doi.org/10.11896/jsjcx.221000056>

### [基于核心句的端到端事件共指消解](#)

End-to-End Event Coreference Resolution Based on Core Sentence

计算机科学, 2023, 50(11): 185-191. <https://doi.org/10.11896/jsjcx.221000078>

### [基于非关键掩码和注意力机制的深度伪造人脸篡改视频检测方法](#)

Deepfake Face Tampering Video Detection Method Based on Non-critical Masks and Attention Mechanism

计算机科学, 2023, 50(11): 160-167. <https://doi.org/10.11896/jsjcx.221100109>

### [路网拓扑感知的轨迹表示学习方法](#)

Road Network Topology-aware Trajectory Representation Learning

计算机科学, 2023, 50(11): 114-121. <https://doi.org/10.11896/jsjcx.221000058>

### [基于时间感知Transformer的交通流预测方法](#)

Time-aware Transformer for Traffic Flow Forecasting

计算机科学, 2023, 50(11): 88-96. <https://doi.org/10.11896/jsjcx.221000201>

# 基于注意力机制的概念增强认知诊断模型

苑冬雪 孙权森 傅鹏

南京理工大学计算机科学与工程学院 南京 210000

(dxyuan@njust.edu.cn)

**摘要** 认知诊断是智能教育系统中的一个基础问题,旨在评估学生对不同知识概念的掌握程度。虽然目前基于深度学习的认知诊断方法相比传统方法有了较大改进,但是其无法充分利用概念之间的潜在相关性。为此,提出一种基于注意力机制的概念增强认知诊断模型(ACECD),通过建模相关概念之间的关系来获得更准确的认知诊断结果。具体来说,首先将学生、练习和概念投影到因子向量来执行复杂交互;然后把概念因子输入自注意力网络中捕获概念之间存在的隐式相关性关系,并用捕获到的隐式关系增强概念因子向量;最后把增强过的概念因子与学生因子和练习因子进行交互,将交互结果输入诊断模块得到最终诊断结果。此外,利用练习因子与概念因子之间的交互修正人为标定Q矩阵的误差。在两个真实世界数据集上与其他方法进行比较,实验结果表明基于注意力机制的概念增强认知诊断模型有效地改善了诊断结果。

**关键词:** 注意力;认知诊断;神经网络

**中图分类号** TP391.4

## Attention Based Concept Enhanced Cognitive Diagnosis

YUAN Dongxue, SUN Quansen and FU Peng

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210000, China

**Abstract** Cognitive diagnosis is a fundamental problem in intelligent education systems, which aims to evaluate the mastery levels of students on different knowledge concepts. Although the performance current deep learning-based cognitive diagnostic methods has improved greatly compared with traditional methods, they cannot fully exploit the potential correlation between concepts. To this end, this paper proposes an attention-based concept enhanced cognitive diagnosis (ACECD) model to obtain more accurate cognitive diagnostic results by modeling the relationship between related concepts. Specifically, we first project students, exercises, and concepts to factor vectors to perform complex interactions, and then feed the concept factors into a self-attention network to capture the implicit correlations that exist between concepts, and concept factor vector can be enhanced with the captured implicit relation. Finally, the enhanced concept factors are interacted with the student factor and the practice factor, and the interacted results are input into the diagnosis module to get the final diagnosis result. In addition, we also use the interaction between the practice factor and the concept factor to correct the bias of the manually-labeled Q matrix. The proposed model is compared with other methods on two real-world datasets, and the experimental results show that the ACECD model effectively improves the diagnostic results.

**Keywords** Attention, Cognitive diagnosis, Neural network

## 1 引言

认知诊断(Cognitive Diagnosis)现已经成为智能教育系统中的一项重要任务,最早起源于教育测量和认知心理学<sup>[1]</sup>。虽然学生的知识状态是不可直接观测的,但是认知诊断可以通过学生的练习记录捕获其知识状态,即对知识概念的掌握程度<sup>[2]</sup>。具体来说,如图1所示,橙色代表学生,蓝色代表练习,绿色代表知识概念。学生练习习题后留下答题记录,根据这些答题记录可推断出他们对这些习题包含的知识概念(比如数学中的三角函数、通分等)的掌握程度,最终得到诊断报告,也就是该名同学对各个知识概念的掌握情况的雷达图。这一诊断结果将会被进一步

应用到练习推荐<sup>[3-4]</sup>及个性化学习<sup>[5-6]</sup>中。

现有的对认知诊断的研究方法主要包括两类:一类为基于统计的传统方法;另一类为基于神经网络的方法。传统方法中,项目响应理论(Item Response Theory, IRT)<sup>[7]</sup>、多维项目响应理论(Multi-dimensional Item Response Theory, MIRT)<sup>[8-9]</sup>,以及确定性输入噪声与门模型(Deterministic Inputs, Noisy and Gate Model, DINA)<sup>[10]</sup>、矩阵分解(Matrix Factorization, MF)<sup>[11]</sup>等,需要依赖手工设计的线性交互函数(逻辑函数、内积等),通过参数估计(最大似然估计等)得到结果,但是这可能不足以捕捉学生和练习之间复杂的非线性关系,并且需要专业人士手工设计交互函数,因此这项工作劳动密集型的。

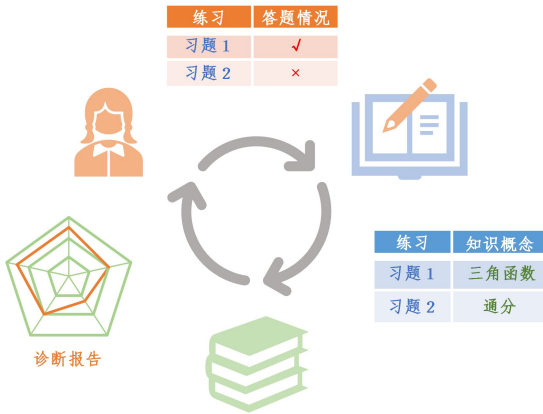


图 1 认知诊断示例(电子版为彩图)

Fig. 1 Example of cognitive diagnosis

为了解决上述问题, NCD (Neural Cognitive Diagnosis) [12] 被提出, 它可以自动地在海量数据中进行认知诊断, 并具备一定的可解释性。尽管如此, 认知诊断领域仍存在很多挑战。

(1) 研究表明概念之间存在相互依赖的关系 [13-14]。具体来说, 如果学生正确或错误地回答了一个练习, 那么他/她的知识掌握程度不仅会受到所回答练习包含的概念的影响, 还会受到其他相关概念的影响。这些概念之间蕴含的潜在隐藏关系包含丰富的信息, 其对提升认知诊断结果的准确性大有裨益。然而现有的方法无法自动挖掘概念之间潜在的隐藏关系, 手动标记这些概念关系将耗费大量人力, 因此迫切需要研究自动学习概念之间关系的方法。其次, 概念之间的影响程度也是存在巨大差异的, 比如, 一些概念与其前提概念(需具备前提概念知识才能掌握后续概念的一对概念)高度相关, 而与其他概念的相关性较低; 另外一些概念与其相似概念(属于同一知识点下的一对概念)高度相关, 而与其他概念相关性较低。衡量这些差异, 建模概念之间的影响系数, 并不是一项普遍的任务。

(2) 在认知诊断中被用来表示练习与知识概念相关性的 Q 矩阵 [10] 通常由教育专家进行标注, 因此可能存在一定偏差。在 NCD [12] 中通过增加文本信息对 Q 矩阵进行完善, 但是这对数据集的要求较高, 需要其提供练习文本, 且模型复杂性较高, 需要通过卷积操作提取练习文本中的概念信息。如何在不增加信息的情况下, 以一种简单的方式自动修正 Q 矩阵, 值得进一步探索。

本文提出了一个基于注意力机制的概念增强认知诊断模型 (Attention-based Concept Enhanced Cognitive Diagnosis Model, ACECD) 来解决上述问题, 以期通过自动挖掘概念之间的差异影响来提高认知诊断的准确性。具体来说, 首先采用对学生因子、练习因子、概念因子分别进行建模的方式, 而不是以往那样直接建模学生能力(学生对知识概念的掌握程度)及练习难度。这样不仅可以充分利用学生与概念之间的关系、练习与概念之间的关系, 还能在得到更高维的概念因子后通过对概念因子应用自注意力机制来自动挖掘概念之间具有差异性的隐藏关系, 以达到增强概念因子的目的。此外, 还通过 Q 矩阵对练习与概念的交互结果进行监督, 修正 Q 矩阵

自身。最后, 使用神经网络对捕捉到的学生对知识概念的掌握程度以及练习难度进行认知诊断。

本文的主要贡献如下:

- (1) 探索了自动捕捉概念之间差异化的潜在关系的方法, 这有助于提高认知诊断的准确性;
- (2) 提出了一个基于注意力机制的概念增强认知诊断模型 (ACECD), 其具有良好的可解释性;
- (3) 在两个真实数据集上进行了大量实验, 并通过消融实验验证了模型的有效性。

## 2 相关工作

### 2.1 传统认知诊断

认知诊断的目标是在预测学生练习回答的过程中诊断学生对知识概念的掌握程度。现有的认知诊断模型中, 最流行的模型为来自教育心理学领域的 DINA [10, 15] 以及 IRT [7]。具体来说, 在 DINA 中, 学生只有掌握了一个练习中的所有知识概念, 才能正确回答该练习, 否则只能通过猜测来回答。给定学生知识状态  $\theta_i$  以及 Q 矩阵中的练习特征向量  $Q_{ij}$  (均为二元), 则可计算代表学生  $s_i$  在有 K 个知识概念的练习  $e_j$  上的潜在特征  $n_{ij}$ 。

$$n_{ij} = \prod_{l=1}^K \theta_i^{Q_{il}} \quad (1)$$

那么学生  $s_i$  在潜在特征为  $\theta_i$  时正确回答练习  $e_j$  的概率  $P(r_{ij} = 1 | \theta_i)$  为:

$$P(r_{ij} = 1 | \theta_i) = c_j^{1-n_{ij}} (1-\xi_j)^{n_{ij}} \quad (2)$$

其中, 练习因子  $c_j$  为练习  $e_j$  的猜测参数, 练习因子  $\xi_j$  为练习  $e_j$  的跳过参数。

在 IRT 中, 学生  $s_i$  的潜在特征为  $\theta_i$  时, 其在练习  $e_j$  上留下正确分数  $r_{ij}$  的概率为:

$$P(r_{ij} = 1 | \theta_i) = \frac{1}{1 + e^{-\alpha_j(\theta_i - \beta_j)}} \quad (3)$$

其中,  $\alpha_j$  代表练习  $e_j$  的区分度,  $\beta_j$  代表练习  $e_j$  的难度,  $\theta_i$  代表学生  $s_i$  的能力, 后两者均为一维且连续的潜在特征, 交互方式为逻辑函数。矩阵分解 MF [11] 与 IRT 一样, 也是使用线性函数; MIRT [8-9] 则是对 IRT [7] 的改进, 从一维增加到多维。它们都依赖于手工设计的线性函数。

### 2.2 基于神经网络的认知诊断

神经网络由于具有强大的拟合能力而在很多领域得到了广泛应用, 比如人脸识别 [16]、语音识别 [17]、自动驾驶 [18] 等等。在教育应用方面, 一些方法 [11, 18] 融合了神经网络, 取得了比传统方法更好的性能。深度知识追踪 (Deep Knowledge Tracing, DKT) [19] 使用递归神经网络 (Recurrent Neural Network) [20] 对学生成绩进行预测, 但其潜在特征无法与知识概念一一对应, 因此不具有可解释性。神经认知诊断 (Neural Cognitive Diagnosis, NCD) [11] 中使用神经网络学习学生和练习之间的复杂交互, 对学生能力和练习难度进行建模。除此之外, Tsutsumi 等 [21] 将深度学习与 IRT 结合, 建立了一个学习者网络和一个项目网络, 实现了更好的特征表示。Tong 等 [22] 提出的增量认知诊断 (Incremental Cognitive Diagnosis, ICD), 通过深度特质网络 (Deep Trait Net-

work,DTN),以归纳的方式获取特征参数,来适应智能教育在线场景的认知诊断。Huang等<sup>[23]</sup>提出了一种基于多任务的群体水平认知诊断(Multi-Task based Group-Level Cognitive Diagnosis, MGCD)框架,合理地建模了不同学习情境下的群体能力与个体能力之间的关系。Zhou等<sup>[24]</sup>通过分析教育背景的特征,提出了一个两阶段的教育背景感知认知诊断(Educational Context-aware Cognitive Diagnosis, ECD)框架。但这些方法均没有考虑概念之间的关系。在认知诊断过程中增加概念关系的建模可能会带来模型性能的提升,因为概念之间的潜在关系有助于更全面且准确地推断出学生对某一概念的掌握程度,从而获得更准确的认知诊断结果。为此,我们提出了 ACECD——一种基于注意力机制的概念增强认知诊断模型,来自动挖掘概念之间存在的潜在差异化关系。在真实数据集上的实验表明,该模型具有更高的准确度,并保证了一定的可解释性。

### 2.3 自注意力机制

注意力最早被计算机视觉领域<sup>[25-26]</sup>所使用。近年来,注意力机制(Attention Mechanism)在自然语言处理领域被提出,并被广泛应用于基于深度学习的自然语言处理任务中。2015年,Bahdanau等<sup>[27]</sup>在机器翻译模型中实现了注意力,并对其进行了改进。在此之后,注意力迅速在各种任务中流行起来。2015年,Xu等<sup>[28]</sup>提出了一个超越传统 Seq2Seq 架构的注意力框架。2016年,Yang等<sup>[29]</sup>通过他们的分层注意力网络证明了注意力可以在各个层面上得到有效利用。2017年 Vaswani 等<sup>[30]</sup>提出的完全由注意力机制搭建而成的 Transformer 网络取得了非常好的性能,证明了注意力机制足以构建最先进的模型。此外,注意力还能将某种类型的解释引入不具备解释性的神经网络模型<sup>[28]</sup>中,这些优点使得注意力在深度学习中广泛流行。

Transformer 启发了许多研究人员,促使了基于自注意力

机制模型的开发,例如 2019 年 Devlin 等<sup>[31]</sup>提出的来自 Transformers 的双向编码器表示。自注意力机制(Self-attention Mechanism)是注意力机制的一个变种,被用来捕获输入样本之间的隐式依赖关系,而不需要外部信息的监督或引导。在认知诊断任务中,概念编码是一个无序的集合,并且不同概念之间的潜在相关性可以被表示为一个完全图。因此,本文将采用自注意力机制来挖掘它们的潜在关系。

## 3 ACECD 模型

### 3.1 任务概述

假设在认知诊断任务中存在一个学生系统,其包含的  $N$  个学生、 $M$  个练习和  $K$  个知识概念分别表示为  $S = \{s_1, s_2, s_3, \dots, s_N\}$ ,  $E = \{e_1, e_2, e_3, \dots, e_M\}$ ,  $C = \{c_1, c_2, c_3, \dots, c_K\}$ 。  $r_{ij}$  代表学生  $s_i$  在练习  $e_j$  上的得分,则学生  $s_i$ 、练习  $e_j$  和得分  $r_{ij}$  可构成一条日志  $R$ 。每个学生可以选择不同的练习进行交互,而每个练习也包含一定的知识概念。  $Q$  矩阵为人工标记的知识相关矩阵,可表示为  $Q = \{Q_{ij}\}_{M \times K}$ ,当练习  $e_i$  包含知识概念  $c_j$  时  $Q_{ij} = 1$ ,否则  $Q_{ij} = 0$ 。

问题定义:输入日志  $R$  以及  $Q$  矩阵,认知诊断将会通过学生的历史答题记录预测学生在未作答练习上的表现,并在这个过程中评估学生对知识概念的掌握程度。

### 3.2 ACECD 框架

ACECD 模型框架如图 2 所示,它包含嵌入层、概念增强层、诊断层这 3 个部分。在嵌入层中,编码学生、概念、练习为  $d$  维的学生、概念、练习因子,同时用练习的 one-hot 向量编码练习区分度。在概念增强层,使用自注意力机制对  $d$  维的概念因子进行加权,自动地学习概念与概念之间的关系。在诊断层,用学生因子、练习因子、概念因子推断学生对知识概念的掌握程度并对  $Q$  矩阵进行修正,最后通过诊断函数预测学生在练习上的得分。

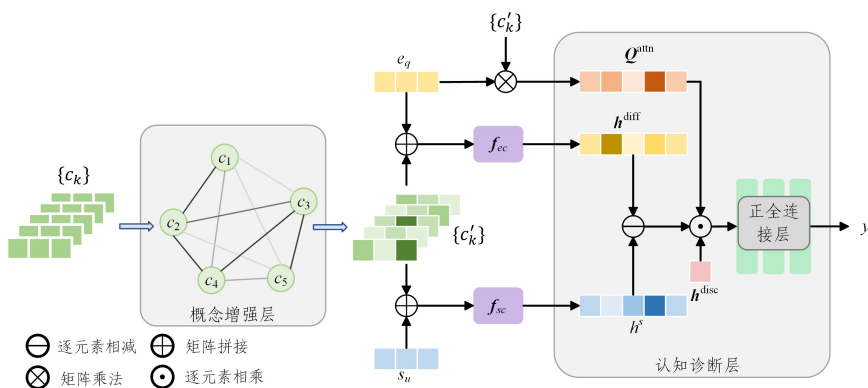


图 2 ACECD 框架

Fig. 2 Framework of ACECD

#### 3.2.1 嵌入层

在嵌入层中,首先对学生、练习、概念分别进行嵌入,得到学生因子、练习因子、概念因子。具体来说,通过输入的学生、练习、概念的 one-hot 向量与可训练矩阵  $W_s$ 、 $W_e$  和  $W_c$  相乘,获得输入所对应的  $d$  维特征向量,其中  $W_s \in \mathbb{R}^{N \times d}$ ,  $W_e \in \mathbb{R}^{M \times d}$ ,  $W_c \in \mathbb{R}^{K \times d}$ 。学生因子  $s_u$ 、练习因子  $e_q$ 、概念因子  $c_k$  可以利用如下公式获得:

$$s_u = \mathbf{x}_u \times W_s, e_q = \mathbf{x}_q \times W_e, c_k = \mathbf{x}_k \times W_c \quad (4)$$

其中,  $s_u, e_q, c_k \in \mathbb{R}^d$ ,  $\mathbf{x}_u \in \{0, 1\}^{1 \times N}$ ,  $\mathbf{x}_q \in \{0, 1\}^{1 \times M}$ ,  $\mathbf{x}_k \in \{0, 1\}^{1 \times K}$ 。

其次,由于同一个概念会对应多个不同难度的练习,因此对每道练习编码了练习区分度因子  $h^{\text{disc}} \in [0, 1]$ ,它代表练习区分能力强学生和能力强学生的程度。  $h^{\text{disc}}$  由以下公式计算:

$$h^{\text{disc}} = \text{sigmoid}(\mathbf{x}_q \times \mathbf{D}) \quad (5)$$

其中,  $\mathbf{D} \in \mathbb{R}^{M \times 1}$  为可训练矩阵。

### 3.2.2 概念增强层

直观上来说,概念及概念之间的关系可以被建模成图结构。因此,我们将概念首先组织成一个完全图,然后利用特征之间的相关性来指示概念之间的隐式关系。在充足的训练数据下,网络可以自适应地学习到这一相关性。在这一层中,通过把自注意力机制作用于概念因子的  $d$  维特征上来学习概念之间可能存在的隐式关系。具体来说,  $K$  个概念因子  $\mathbf{C} = \{c_k | k=1, 2, \dots, K\}$ ,  $\mathbf{C} \in \mathbb{R}^{K \times d}$  首先通过线性变换得到  $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{K \times d}$ :

$$\mathbf{Q} = \mathbf{C}\mathbf{W}_q, \mathbf{K} = \mathbf{C}\mathbf{W}_k, \mathbf{V} = \mathbf{C}\mathbf{W}_v \quad (6)$$

其中,  $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d \times d}$  都是可学习的权重。

然后利用概念因子之间的特征相关性来计算概念相关权重  $\mathbf{A} \in \mathbb{R}^{K \times K}$ :

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d}}\right) \quad (7)$$

通过数据驱动的方式,在网络得到充分训练之后,  $\mathbf{A}$  可以捕获到概念之间的隐含关系。

最后,利用得到的概念相关权重来聚合相关概念,从而得到通过概念关系增强过的概念因子  $\mathbf{C}' \in \mathbb{R}^{K \times d}$ :

$$\mathbf{C}' = \text{sigmoid}(\mathbf{A}\mathbf{V}) \quad (8)$$

通过隐式关系增强过的概念因子包含了从相邻概念之间聚合的信息,有利于诊断层得到更准确的诊断结果。

### 3.2.3 诊断层

在诊断层中,通过全连接层对学生因子、练习因子与包含概念关系的概念因子进行融合,得到代表学生对  $K$  个知识概念的掌握程度向量和练习在  $K$  个知识概念上的难度向量。

$$\mathbf{h}^s = f_{\sigma}(s_u, \mathbf{C}') = \text{sigmoid}(\mathbf{W}_{\sigma} \times (s_u \oplus \mathbf{C}') + b_{\sigma}) \quad (9)$$

其中,  $\mathbf{h}^s \in \mathbb{R}^{1 \times K}$  为学生能力向量,  $\mathbf{W}_{\sigma} \in \mathbb{R}^{2d \times 1}$ ,  $\oplus$  代表拼接操作。

$$\mathbf{h}^{\text{diff}} = f_{\sigma}(e_q, \mathbf{C}') = \text{sigmoid}(\mathbf{W}_{\sigma} \times (e_q \oplus \mathbf{C}') + b_{\sigma}) \quad (10)$$

其中,  $\mathbf{h}^{\text{diff}} \in \mathbb{R}^{1 \times k}$  为练习难度向量,  $\mathbf{W}_{\sigma} \in \mathbb{R}^{2d \times 1}$ 。

进一步,在诊断层预测学生  $s_u$  在练习  $e_q$  上的表现:

$$\mathbf{x} = \mathbf{Q}_q \odot (\mathbf{h}^s - \mathbf{h}^{\text{diff}}) \times h^{\text{disc}} \quad (11)$$

$$\mathbf{Q}_q = \mathbf{x}_q \times \mathbf{Q} \quad (12)$$

其中,  $\mathbf{x} \in \mathbb{R}^{1 \times K}$ ,  $\mathbf{Q} \in \mathbb{R}^{M \times K}$  为人工标记并预先给出的  $\mathbf{Q}$  矩阵,它与练习 one-hot 相乘,得到练习  $e_q$  的知识相关向量  $\mathbf{Q}_q \in \{0, 1\}^{1 \times K}$ ;  $\odot$  代表对应元素相乘。

但是,由于专家手动标记的  $\mathbf{Q}$  矩阵可能存在一定误差,为此我们将探索一种使  $\mathbf{Q}$  矩阵更为准确且自适应的方法。首先,通过练习因子与概念因子交互得到可学习的练习概念相关向量,并使用  $\mathbf{Q}$  矩阵对其进行监督学习,虽然  $\mathbf{Q}$  矩阵存在一定误差,但我们认为它仍然具有较高的参照价值。接下来,将使用这里的练习概念相关向量代替人工标记的  $\mathbf{Q}$  矩阵。具体做法如下:

$$\mathbf{x} = \mathbf{Q}_q' \odot (\mathbf{h}^s - \mathbf{h}^{\text{diff}}) \times h^{\text{disc}} \quad (13)$$

$$\mathbf{Q}_q' = \text{softmax}(e_q \times \mathbf{C}'^T) \quad (14)$$

其中,  $e_q \in \mathbb{R}^{1 \times d}$ ,  $\mathbf{C}' \in \mathbb{R}^{K \times d}$ ,  $\mathbf{Q}_q' \in \mathbb{R}^{1 \times K}$ 。

然后,对  $\mathbf{x}$  应用三层全连接,得到模型对学生在练习上的

得分预测  $y$ :

$$f_1 = \phi(\mathbf{W}_1 \times \mathbf{x}^T + b_1) \quad (15)$$

$$f_2 = \phi(\mathbf{W}_2 \times f_1 + b_2) \quad (16)$$

$$y = \phi(\mathbf{W}_3 \times f_2 + b_3) \quad (17)$$

其中,  $\phi$  为激活函数,这里使用 sigmoid。

为了增加模型的可解释性,应用 IRT<sup>[7]</sup> 中常用的单调性假设。单调性假设认为学生正确回答练习的概率越大,那么其知识掌握程度就越好。为了在训练过程中满足单调性假设,我们采取的策略为限制  $\mathbf{W}_1, \mathbf{W}_2$  和  $\mathbf{W}_3$  的每个元素为正,类似 NCDM<sup>[12]</sup>。

### 3.3 模型训练

通过计算学生  $s_i$  在练习  $e_j$  上的真实得分  $r_{ij}$  与模型预测值  $y_{ij}$  之间的交叉熵来监督网络的训练。

$$\text{loss}_r = - \sum_i \sum_j (r_{ij} \log y_{ij} + (1 - r_{ij}) \log (1 - y_{ij})) \quad (18)$$

此外,手工标注的  $\mathbf{Q}$  矩阵虽然带有偏差,但是依然能为预测的  $\mathbf{Q}'$  提供一定的先验信息。为此,设计一个均方差损失函数让优化的  $\mathbf{Q}'$  能学习到这种先验,并且从练习和概念的交互中修复  $\mathbf{Q}$  矩阵的偏差:

$$\text{loss}_Q = \sum_1^K (\mathbf{Q} - \mathbf{Q}')^2 \quad (19)$$

最终的损失函数设计为以上两个损失函数的加权和:

$$\text{loss}_{\text{ACECD}} = \text{loss}_r + \lambda \text{loss}_Q \quad (20)$$

其中,  $\lambda$  为损失函数的权重,在实验中设置  $\lambda = 1$ 。

经过训练后,  $\mathbf{h}'$  即为得到的诊断结果,代表学生能力也即学生对知识概念的掌握程度。

### 3.4 模型复杂度分析

如算法 1 所示,在每轮迭代中,ACECD 首先在概念增强层中计算所有概念之间的隐式相关性并对概念因子进行增强,这一步的算法复杂度为  $O(K^2)$ ;在诊断层中,计算学生能力向量的复杂度为  $O(NK)$ ,计算练习难度的复杂度为  $O(MK)$ ,计算最终得分预测的复杂度为  $O(NK)$ 。但是由于概念的数量远小于学生和练习的数量,因此概念增强层所带来的额外计算代价并不高。最终,每轮迭代中算法的总的时间复杂度为:  $O((N+M+K)K)$ 。

**算法 1** 基于注意力机制的概念增强认知诊断模型

输入: 学生练习交互日志  $R$ , 知识相关矩阵  $\mathbf{Q}$

输出: 学生在练习上的预测得分  $y$ , 学生能力向量  $\mathbf{h}'$

1. 将学生练习交互日志和练习知识概念相关向量处理为序列
2.  $s_u \leftarrow$  student embedding  
 $e_q \leftarrow$  exercise embedding  
 $c_k \leftarrow$  concept embedding  
/\* 将学生练习概念编码为  $d$  维特征向量 \*/
3.  $\mathbf{Q}, \mathbf{K}, \mathbf{V} \leftarrow \mathbf{C}\mathbf{W}_q, \mathbf{C}\mathbf{W}_k, \mathbf{C}\mathbf{W}_v$  /\* 对  $K$  个概念因子组合而成的  $\mathbf{C}$  进行线性变换 \*/
4.  $\mathbf{A} \leftarrow \text{softmax}((\mathbf{Q} \times \mathbf{K}^T) \times 1/\sqrt{d})$  /\* 通过概念因子之间的特征相关性计算概念相关权重 \*/
5.  $\mathbf{C}' \leftarrow \text{sigmoid}(\mathbf{A}\mathbf{V})$  /\* 通过聚合相邻概念之间信息,增强概念因子 \*/
6.  $\mathbf{h}' \leftarrow \text{sigmoid}(\mathbf{W}_{\sigma} \times (s_u \oplus \mathbf{C}') + b_{\sigma})$  /\* 将学生因子与增强过的概念因子融合,获得学生能力向量 \*/

7.  $\mathbf{h}^{\text{diff}} \leftarrow \text{sigmoid}(\mathbf{W}_{\text{ec}} \times (\mathbf{e}_i \oplus \mathbf{C}') + \mathbf{b}_{\text{ec}}) / *$  将练习因子与增强过的概念因子融合,获得练习难度向量  $*$  /
8.  $\mathbf{Q}_i' \leftarrow \text{softmax}(\mathbf{e}_i \times \mathbf{C}'^T) / *$  通过练习因子与概念因子交互修正 Q 矩阵  $*$  /
9.  $\mathbf{x} \leftarrow \mathbf{Q}_i' \odot (\mathbf{h}^{\text{s}} - \mathbf{h}^{\text{diff}}) \times \mathbf{h}^{\text{disc}} / *$  将学生能力向量与练习难度向量进行交互  $*$  /
10.  $y = \phi(\mathbf{W}_3 \times \phi(\mathbf{W}_2 \times \phi(\mathbf{W}_1 \times \mathbf{x}^T + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3) / *$  通过三层全连接获得学生在练习上的预测得分  $*$  /
11. 通过反向传播更新模型全部参数

## 4 实验

### 4.1 数据集描述

实验中使用了两个来自真实世界的数据集,分别为 ASSIST 和 Junyi。ASSIST(ASSISTments 2009-2010 “skill builder”)是由 ASSISTments 在线辅导系统<sup>[32]</sup>收集的公开数据集,它提供了 4000 余名学生在 17000 多道练习上的 324000 多次回答,其练习涵盖了 123 个知识概念。Junyi 是来自在线教育平台 Junyi 学院的公开数据集,它提供了 10000 个学生在 800 多道练习上的 353000 多次回答,其练习涵盖了 835 个知识概念。由于学生在在线学习系统中可以多次提交答案,原始数据集将不利于静态认知诊断,为此对数据集进行预处理,仅采用学生第一次提交的记录。除此之外,为了保证每个学生有足够的练习数据进行诊断<sup>[33]</sup>,将删除日志不足 15 条的学生记录。训练集和测试集采用了 8:2 的划分方式。表 1 列出了两个数据集的统计信息。

表 1 数据集的统计数据  
Table 1 Statistics of datasets

Statistics	Junyi	ASSIST
Students	10000	4163
Exercises	835	17746
Knowledge concepts	835	123
Response records	353835	324572
Knowledge concepts per exercise	1.00	1.19

### 4.2 实验设置

#### 4.2.1 基线与评估标准

为了验证模型的有效性,选择认知诊断的传统方法与基于神经网络的方法在相同数据集上进行对比实验。

(1)MF 矩阵分解:分解分数矩阵来预测学生表现。

(2)IRT 项目响应理论:通过线性函数模拟一维的学生和练习特征。

(3)MIRT 多维项目响应理论:IRT 的多维扩展,模拟了学生和练习的多维特征。

(4)NCDM 神经认知诊断模型:使用神经网络对高阶和复杂的学生练习交互函数进行建模。

为了衡量模型的性能,采用了分类和回归的指标。

均方根误差(RMSE)<sup>[34]</sup>:量化预测分数和实际分数之间的距离。

预测精度(ACC)<sup>[12]</sup>和 ROC 曲线下面积(AUC)<sup>[35]</sup>:评估模型。

#### 4.2.2 参数设置

全连接层(式(15)一式(17))的维度分别为 512,256,1。

所有层的激活函数均使用 sigmoid。使用 Xavier 来初始化所有网络参数<sup>[36]</sup>,每个参数抽样自  $U(-std, std)$ ,其中  $std = \sqrt{\frac{2}{n_{\text{in}} + n_{\text{out}}}}$ ,  $n_{\text{in}}$  是输入权重的神经元数量,  $n_{\text{out}}$  是输出权重的神经元数量。使用 Adam 算法<sup>[37]</sup>进行梯度更新。批处理大小设置为 128。将学生、概念、练习的嵌入维度设置为 8。

所有模型均由 PyTorch 使用 Python 3.7 实现,所有实验均在配置有频率为 3.20 GHz 的 CPU、RTX 2080Ti GPU 和 16GB 内存的 Linux 服务器上运行。

### 4.3 实验结果

#### 4.3.1 基线性能比较

因为认知诊断模型的结果是学生知识概念的掌握程度,而学生的真实知识状态很难直接获得,所以认知诊断模型的性能通常是根据诊断结果来预测学生表现获得,也就是说预测任务的表现可以间接评估模型。为了评估 ACECD 模型的有效性,在两个真实数据集上将其与基线进行比较。

表 2 列出了所有模型的预测性能和评估指标结果,包括准确性(ACC)、均方根误差(RMSE)、曲线下面积(AUC)。最佳结果用粗体表示。从表中结果可以看出:首先,通过神经网络学习学生与练习之间复杂交互的方式相比传统模型在预测性能上有了大幅提升;其次,融合了概念关系的 ACECD 模型相比 NCD 模型在预测性能上也有了一定程度的提升。因为 ACECD 模型利用注意力机制挖掘了概念之间存在的潜在差异化关系,并在诊断层通过融合了概念间关系的概念因子生成学生能力向量以及练习难度向量来预测学生的练习成绩,这样在预测成绩时模型不仅考虑了练习所包含的知识概念,同时还考虑了与这些概念相关的概念的差异化影响;此外,在诊断层,修正弥补了一些由专家进行标注的 Q 矩阵可能存在的误差,因此 ACECD 模型在预测学生成绩上表现良好。这也证明了,增加概念之间的潜在关系以及概念之间的差异影响,对提升认知诊断结果的准确性有着重要意义。具体来说,相较于最优基线 NCD,在学生成绩预测准确性上,ACECD 模型在数据集 ASSIST 上提升了 0.2%,在 Junyi 上提升了 2.6%,从而验证了该框架的有效性。

表 2 学生成绩预测实验结果

Table 2 Experimental results of student performance prediction

(单位:%)

Methods	ASSIST			Junyi		
	ACC $\uparrow$	AUC $\uparrow$	RMSE $\downarrow$	ACC $\uparrow$	AUC $\uparrow$	RMSE $\downarrow$
IRT	64.26	69.83	46.59	67.60	77.50	42.68
MIRT	71.70	74.94	45.17	75.13	79.89	41.17
MF	67.12	76.45	44.51	68.34	76.44	43.73
NCD	73.14	75.94	43.08	74.43	79.09	41.72
ACECD	<b>73.34</b>	<b>76.98</b>	<b>42.37</b>	<b>77.03</b>	<b>82.45</b>	<b>39.85</b>

#### 4.3.2 消融实验

在两个数据集上运行了 ACECD 模型,并将其与删掉部分模块的模型进行对比,以了解概念增强以及修正 Q 矩阵对模型性能的影响。表 3 中,ACECD+q 代表模型只包含修正 Q 矩阵而不包含概念增强模块,ACECD+ce 代表只包含概念增强模块而不包含修正 Q 矩阵模块。最佳结果用粗体表示。

表3 模型的消融测试结果

Table 3 Results of ablation testing on our model

Methods	ASSIST			Junyi		
	ACC↑	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓
ACECD	<b>73.34</b>	<b>76.98</b>	<b>42.37</b>	<b>77.03</b>	<b>82.45</b>	<b>39.85</b>
ACECD+q	73.20	76.29	42.86	76.02	81.27	41.10
ACECD+ce	73.28	76.73	42.46	76.76	81.94	40.24

(单位:%)

从表3中结果可以看出移除每个模块都将导致预测性能下降,证明了概念增强以及修正Q矩阵模块对提升模型认知诊断精确度至关重要。概念增强模块对模型的优化效果尤为明显,因为概念增强模块融合了由注意力机制挖掘而来的概念关系,这些差异化的概念关系中所包含的丰富信息将帮助诊断层更为准确地预测学生在练习上的成绩;此外,由于专家标注的Q矩阵可能存在一定的人工误差,比如练习相关的概念可能会被遗漏或是误标,而这将对学生在练习上的成绩预测产生影响,通过修正Q矩阵模块可以减少这些误差带来的影响。

具体来说,在数据集ASSIST上,移除概念增强模块导致准确率下降0.14%,移除修正Q矩阵模块导致准确率下降0.06%;在数据集Junyi上,移除概念增强模块导致准确率下降1.01%,移除修正Q矩阵模块导致准确率下降0.27%。

#### 4.3.3 可解释性分析

为了评估模型的可解释性,也就是模型是否合理,在ASSIST数据集上进一步进行了几项实验。通常来讲,若一个学生比另一个学生对知识概念 $k$ 的掌握程度更高,那他将更可能回答对与 $k$ 有关的练习。为此,采用一致度(Degree of Agreement, DOA)<sup>[38]</sup>作为评估模型可解释性的指标。对于知识概念 $k$ , $DOA(k)$ 为:

$$DOA(k) = \frac{1}{Z} \sum_{a=1}^N \sum_{b=1}^N I(F_{ak}^s > F_{bk}^s) \times H \quad (21)$$

$$H = \frac{\sum_{j=1}^M I(Q_{jk} = 1) \wedge J(j, a, b) \wedge I(r_{aj} > r_{bj})}{\sum_{j=1}^M I(Q_{jk} = 1) \wedge J(j, a, b) \wedge I(r_{aj} \neq r_{bj})} \quad (22)$$

$$Z = \sum_{a=1}^N \sum_{b=1}^N \delta(F_{ak}^s, F_{bk}^s) \quad (23)$$

其中, $F_{ak}^s$ 为学生 $a$ 对知识概念 $k$ 的掌握程度。如果学生 $a$ 对知识概念 $k$ 的掌握程度比学生 $b$ 高,则 $I(F_{ak}^s > F_{bk}^s) = 1$ ;否则 $I(F_{ak}^s > F_{bk}^s) = 0$ 。如果练习 $j$ 包含知识概念 $k$ ,则 $I(Q_{jk} = 1) = 1$ ;否则 $I(Q_{jk} = 1) = 0$ 。如果学生 $a$ 和学生 $b$ 都做了练习 $j$ ,则 $J(j, a, b) = 1$ ;否则 $J(j, a, b) = 0$ 。如果学生 $a$ 在练习 $j$ 上的得分比学生 $b$ 高,则 $I(r_{aj} > r_{bj}) = 1$ ;否则 $I(r_{aj} > r_{bj}) = 0$ 。如果学生 $a$ 在练习 $j$ 上的得分不等于学生 $b$ 在练习 $j$ 上的得分,则 $I(r_{aj} \neq r_{bj}) = 1$ ;否则 $I(r_{aj} \neq r_{bj}) = 0$ 。用所有知识概念的DOA( $k$ )的平均值来评估诊断结果的合理性。

在传统模型中DINA的潜在特征与知识概念之间有明显的对应关系,在基于神经网络的模型NCDM中增加了单调性假设,因此它们都具有一定的可解释性。本节分别对比了NCDM,DINA以及随机估计学生对知识概念掌握程度的DOA值来验证ACECD模型的可解释性。实验结果如图3所示,可以看出我们的优化在提高预测精确度的同时也保证了模型的可解释性,因为ACECD模型也采用了类似NCDM

中的单调性假设来确保模型的可解释性。

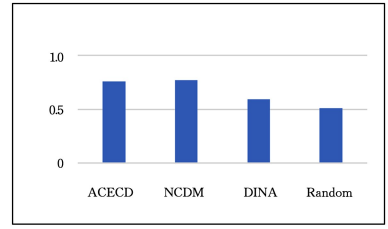


图3 模型的DOA值

Fig. 3 DOA results of models

**结束语** 目前传统的认知诊断方法和基于神经网络的认知诊断方法都没有自动建模概念之间的隐式依赖关系,而探索概念之间隐藏关系所包含的丰富信息对提升基于神经网络的认知诊断的性能有着重要意义,因此本文提出了一种基于注意力的概念增强认知诊断模型ACECD。ACECD使用自注意力机制充分挖掘概念之间的潜在差异化关系,并利用它们之间的关系对概念因子进行增强,可以更全面地对学生的概念掌握程度进行诊断。此外,由于手工标注的Q矩阵可能存在偏差,ACECD还将练习与增强的特征进行交互,从而修正Q矩阵的偏差。在两个世界真实数据集上的结果证明了ACECD模型的有效性,同时ACECD也保证了认知诊断方法的可解释性。

#### 参考文献

- [1] CARBONELL J R. AI in CAI: An artificial-intelligence approach to computer-assisted instruction[J]. IEEE Transactions on Man Machine Systems, 1971, 11(4): 190-202.
- [2] LIU Q, WU R, CHEN E, et al. Fuzzy cognitive diagnosis for modelling examinee performance[J]. ACM Transactions on Intelligent Systems and Technology, 2018, 9(4): 1-26.
- [3] TANG X, CHEN Y, LI X, et al. A reinforcement learning approach to personalized learning recommendation systems[J]. British Journal of Mathematical and Statistical Psychology, 2019, 72(1): 108-135.
- [4] ZHOU Y, HUANG C, HU Q, et al. Personalized learning full-path recommendation model based on LSTM neural networks[J]. Information Sciences, 2018, 444: 135-152.
- [5] LIU Q, TONG S, LIU C, et al. Exploiting cognitive structure for adaptive learning[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019: 627-635.
- [6] ELLIOTT M, PINAR W F, REYNOLDS W M, et al. Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses[J]. Brock Education A Journal of Educational Research and Exercise, 2010, 13(1): 100.
- [7] EMBRETSON S E, REISE S P. Item response theory[M]. Psychology Press, 2013.
- [8] ACKERMAN T A, GIERL M J, WALKER C M. Using multidimensional item response theory to evaluate educational and psychological tests[J]. Educational Measurement: Issues and Practice, 2003, 22(3): 37-51.
- [9] RECKASE M D. Multidimensional item response theory models[M]. New York: Springer, 2009: 79-112.

- [10] DE LA TORRE J. DINA model and parameter estimation: A didactic[J]. *Journal of Educational and Behavioral Statistics*, 2009, 34(1):115-130.
- [11] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. *Computer*, 2009, 42(8):30-37.
- [12] WANG F, LIU Q, CHEN E, et al. Neural cognitive diagnosis for intelligent education systems[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020:6153-6161.
- [13] ELLIS H C. *The transfer of learning*[M]. Macmillan, 1965.
- [14] WOODWORTH R S, THORNDIKE E L. The influence of improvement in one mental function upon the efficiency of other functions[J]. *Psychological Review*, 1901, 8(3):247.
- [15] VON DAVIER M. The DINA model as a constrained general diagnostic model: Two variants of a model equivalency[J]. *British Journal of Mathematical and Statistical Psychology*, 2014, 67(1):49-71.
- [16] LAWRENCE S, GILES C L, TSOI A C, et al. Face recognition: A convolutional neural-network approach[J]. *IEEE Transactions on Neural Networks*, 1997, 8(1):98-113.
- [17] CHAN W, JAITLEY N, LE Q, et al. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition[C]// *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016:4960-4964.
- [18] BOJARSKI M, DEL TESTA D, DWORAKOWSKI D, et al. End to end learning for self-driving cars[J]. *arXiv:1604.07316*, 2016.
- [19] PIECH C, BASSEN J, HUANG J, et al. Deep knowledge tracing[J]. *Advances in Neural Information Processing Systems*, 2015, 28:1-9.
- [20] WILLIAMS R J, ZIPSER D. A learning algorithm for continually running fully recurrent neural networks[J]. *Neural Computation*, 1998, 1(2):270-280.
- [21] TSUTSUMI E, KINOSHITA R, UENO M. Deep-IRT with independent student and item networks[J]. *International Educational Data Mining Society*, 2021:510-517.
- [22] TONG S, LIU J, HONG Y, et al. Incremental cognitive diagnosis for intelligent education[C]// *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2022:1760-1770.
- [23] HUANG J, LIU Q, WANG F, et al. Group-level cognitive diagnosis: A multi-task learning perspective[C]// *2021 IEEE International Conference on Data Mining*. IEEE, 2021:210-219.
- [24] ZHOU Y, LIU Q, WU J, et al. Modeling context-aware features for cognitive diagnosis in student learning[C]// *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021:2420-2428.
- [25] LAROCHELLE H, HINTON G E. Learning to combine foveal glimpses with a third-order boltzmann machine[J]. *Advances in Neural Information Processing Systems*, 2010, 23:1-9.
- [26] MNH V, HEES N, GRAVES A. Recurrent models of visual attention[J]. *Advances in Neural Information Processing Systems*, 2014, 27:1-9.
- [27] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]// *International Conference on Learning Representations*. 2014.
- [28] XU K, BA J, KIRO S R, et al. Show, attend, and tell: Neural image caption generation with visual attention[C]// *International Conference on Machine Learning*. PMLR, 2015:2048-2057.
- [29] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification[C]// *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies*. 2016:1480-1489.
- [30] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30:1-11.
- [31] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]// *Proceedings of NAACL-HLT*. 2019:4171-4186.
- [32] 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.
- [33] PANDEY S, SRIVASTAVA J. RKT: Relation-aware self-attention for knowledge tracing[C]// *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020:1205-1214.
- [34] PEI H, YANG B, LIU J, et al. Group sparse bayesian learning for active surveillance on epidemic dynamics[C]// *Thirty-Second AAAI Conference on Artificial Intelligence*. 2017.
- [35] BRADLEY A P. The use of the area under the ROC curve in the evaluation of machine learning algorithms[J]. *Pattern Recognition*, 1997, 30(7):1145-1159.
- [36] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]// *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. 2010:249-256.
- [37] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. *arXiv:1412.6980*, 2014.
- [38] FOUSS F, PIROTTE A, RENDERS J M, et al. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(3):355-369.



**YUAN Dongxue**, born in 1996, postgraduate. Her main research interests include educational data mining and so on.



**SUN Quansen**, born in 1963, Ph.D., professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include image recognition and computer vision.