

# 一个总体框架 用于认知诊断

Wang Fei, Liu Qi, IEEE 成员, Chen Enhong, IEEE 资深成员, Huang Zhenya, Yin Yu, Wang Shijin, Su Yu

认知诊断广泛应用于需要评估用户认知状态的场景,如游戏、临床测量等。特别是在近几十年来日益流行的智能教育中,认知诊断作为发现学生对特定知识概念熟练程度的基本模块。现有的方法通常是通过人工设计函数(如 logistic 函数)来挖掘学生运动过程的线性交互。然而,学生与练习之间的认知交互是一个复杂的过程,过度的简化会导致拟合不足,从而得到不准确的诊断结果。此外,手工设计的交互功能相对不灵活,限制了其可扩展性。这导致除了反应日志外,认知过程中缺乏对有用的非数值信息的考虑。在本文中,我们提出了一个通用的神经认知诊断(NeuralCD)框架以及几个实现的模型(一个基本实现的 NeuralCDM 和三个扩展),在这里,我们将学生和练习投射到因子向量中,并结合神经网络来学习复杂的练习交互作用。为了保证诊断结果的可解释性,这对认知诊断至关重要,我们在 NeuralCD 框架中引入了单调性假设。此外,NeuralCD 是一个通用的框架,具有良好的可扩展性。我们通过证明 NeuralCD 可以覆盖一些传统的模型来说明它的通用性。然后,我们论证了 NeuralCD 的可扩展性,这有利于未来的发展。一方面,我们演示了基于内容的扩展,提供了探索练习文本(CNCD-Q 和 CNCD-F)的丰富内容的示例。另一方面,我们展示了一个基于知识关联的扩展,以表明 NeuralCD 对于结构调整是灵活的,从而解决特定的问题。例如,我们通过将 NeuralCD 扩展为知识关联考虑(KaNCd),改进了对学生未揭示知识概念的诊断结果。在真实数据集上的大量实验结果显示了 NeuralCD 框架在准确性和可解释性方面的有效性。

指数术语——智能教育、个性化学习、认知诊断、神经网络

## 1 引言

认知诊断的研究已经进行了几十年,研究人员(尤其是心理测量学和教育学的研究人员)取得了丰富的成果。认知诊断的目的是从一个人的行为(如测试)中发现一个人的认知状态(如技能熟练程度)

?中国科学技术大学计算机科学与技术学院大数据分析与应用安徽省重点实验室,安徽合肥 230052;认知智能国家重点实验室,安徽合肥 230088  
E-mail: {wf314159, yxonic}@mail.ustc.edu.cn, {qiliuql, cheneh, huangzhy}@ustc.edu.cn。  
?王世进就职于科大讯飞人工智能研究中心,湖北武汉 430058;同时就职于认知智能国家重点实验室,安徽合肥 230088。电子邮件:sjwang3@iflytek.com。  
?苏宇就职于合肥师范大学,安徽合肥 230601;国家合肥综合科学中心人工智能研究所,安徽合肥 230088。电子邮件:yusu@hfnu.edu.cn。

稿件于 2021 年 11 月 30 日收到;2022 年 6 月 23 日修订;已于 2022 年 8 月 6 日接受。发布日期:2022 年 8 月 23 日;2023 年 6 月 21 日。  
国家自然科学基金项目(no. 61922073, no. U20A20229, no. 62106244);中央高校基本科研业务费专项资金(no. WK2150110021);科大科大联合研究计划(no. Iflytek Joint Research Program)。关键词:岩石力学,蠕变,应力-应变关系,数值模拟  
(通讯作者:刘琦,陈恩红。)  
由 Z. Wang 推荐接受。  
数字对象标识符编号 10.1109/tkde.2022.3201037

结果)。在现实世界的许多场景中,如游戏[1],临床测量[2],[3]和教育,这是一个必要的和基本的任务,在这些场景中,用户(如玩家,患者,学生)的能力需要评估,因此受到广泛的关注。具体来说,在智能教育系统[4],[5]中,认知诊断的目的是发现学生在学习过程中的状态,如他们对特定知识概念[6]的熟练程度。图 1 展示了一个认知诊断的玩具例子。一般来说,学生通常首先选择练习一套练习题(如: $e_1, \dots, e_4$ ),并留下他们的回答(例如,对或错)。然后,我们的目标是根据相应的概念(如方程式)推断出它们的实际知识状态。在实践中,这些诊断报告是必要的,因为它们为进一步支持的基础,如运动建议,有针对性的训练[7]和计算机化的适应性测试[8]。

在心理测量学领域,人们对认知诊断进行了大量的研究,如确定性输入、噪声和门模型(DINA)[9]、项目反应理论(IRT)[10]和多维 IRT (MIRT)[11]。尽管取得了一些效果,这些作品依赖于手工制作的功能,模拟学生和问题之间的互动。交互函数的设计是基于对实际交互过程进行简化的假设,且大多是线性的,如 IRT[10]中的类逻辑函数或矩阵分解[12]中的内积。然而,两者之间的互动

1041-4347©2022 IEEE. 允许个人使用,但是重新发布/再发布需要 IEEE 的许可。请参阅 <https://www.ieee.org/publications/rights/index.html> 了解更多信息。

授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。

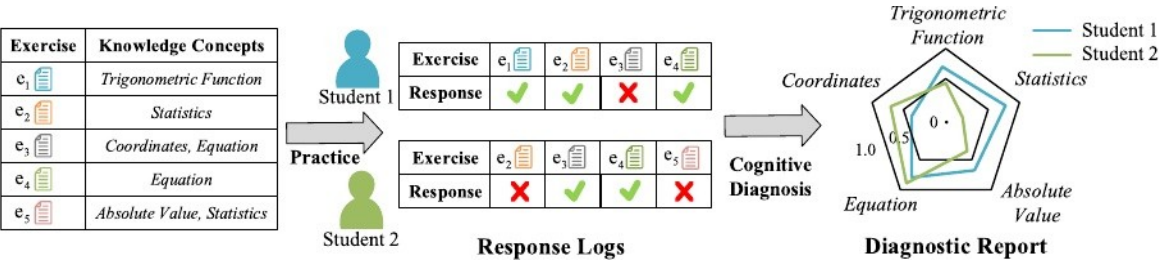


图 1。认知诊断的一个玩具例子。

学生与习题是一个复杂的非线性过程，过度简化会导致过程拟合不足，从而得到不准确的诊断结果，限制了模型的适用范围。此外，这些作品主要是针对基于量表的测试，即对一组考生进行同样的一组小问题的测试，如学校的期末考试。在基于量表的测试中，所有的考生都应该回答所有的问题，因此回答数据是完整的，通常不是很大。而对于认知诊断的更广泛应用，数据可以通过离线考试和在线自主学习等不同场景收集。因此，响应数据可以体积大但分布稀疏，可以使用更多类型的数据(如文本、图形)。对于简单的函数，数据背后的交互模式可能更加难以捉摸，这使得手工设计交互函数变得不切实际。另一方面，幸运的是，数据的积累为我们提供了一个应用数据驱动的方法来发现复杂交互函数[13]的机会。

在本文中，我们以一种有原则的方式提出了一个神经认知诊断(NeuralCD)框架来解决这个问题，将神经网络整合到复杂的非线性交互模型中。虽然神经网络对连续函数的逼近能力已经在很多领域得到了证明，如自然语言处理[14]和推荐系统[15]，但由于以下领域的挑战，神经网络对认知诊断的适应仍然是非常重要的。首先，诊断结果的可解释性，如让学生掌握某些知识概念(如方程式)对认知诊断至关重要。然而，神经网络的黑盒特性使得它们很难得到这样可以解释的结果。第二，响应日志中包含的信息对于认知诊断来说并不完整。额外的资源，如练习文本内容，有价值的信息(如阅读理解的难度)，有利于认知诊断。因此，有必要确保我们提议的框架是可扩展的，以便从这些资源中聚合信息。第三，一个广泛适用的框架应该具有可灵活扩展的结构，以满足不同情况下的不同需求。例如，在图 1 中，学生 1 没有回答与绝对值相关的习题，学生 2 没有回答与三角函数相关的习题。这是一种常见的现象，特别是在没有按比例进行的在线练习中，由于学生所做的知识概念总量大，问题数量有限，所以在学生的回答日志中，知识概念的覆盖并不完整。在这种情况下，诊断

当某些知识概念没有出现在学生的反应历史中时，模型需要处理知识覆盖问题，以获得可靠的诊断结果。

我们提出一个 NeuralCD 框架来应对这些挑战。我们首先介绍如何使用 NeuralCD 来近似学生和练习之间的交互以获得准确和可解释的诊断结果。正如我们在前期工作[16]中提出的，在 NeuralCD 中，我们将学生和练习题投影到因子向量中，并利用多层方法对练习题回答过程中的复杂交互进行建模。为了保证模型的可解释性，使其无法提供可理解的诊断结果，从而成为纯粹的预测模型，我们在多层面上采用了教育性[11]的单调性假设。然后，我们提出了一个基本的框架实现，称为 NeuralCDM，在此框架中，我们简单地从传统的  $q$  矩阵中提取运动因子向量，实现了非负全连接层的单调性，证明了该框架的可行性。传统的  $Q$  矩阵是一种练习知识关联矩阵，其中  $Q_{ij}$  四分之一如果运动  $e_i$  包含知识概念  $k_j$  否则为 0。图 1 中练习的  $q$  矩阵如图 2 所示。

在此之后，本文进一步讨论和论证了 NeuralCD 的两个额外优势，即通用性和可扩展性。

NeuralCD 框架的通用性在于它能够涵盖一些传统的模型，如 MF、IRT、MIRT。这些模型可以看作是 NeuralCD 特殊病例的退化。通过适当的神经网络结构，NeuralCD 能够自动学习适合数据的不同交互函数。

至于可扩展性，我们强调认知诊断框架应该对额外的信息或更好的结构开放。为此，我们提出了两个模型，分别从基于内容的扩展和基于知识关联的扩展两个方面来论证 NeuralCD 的可扩展性。在基于内容的扩展中，我们演示了如何利用神经网络从练习文本中获取信息来扩展框架在

$k_1$  : Trigonometric Function

$k_2$  : Statistics

$k_3$  : Coordinates

$k_4$  : Equation

$k_5$  : Absolute Value

$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{matrix}$

图二。q 矩阵的一个例子。  
授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。

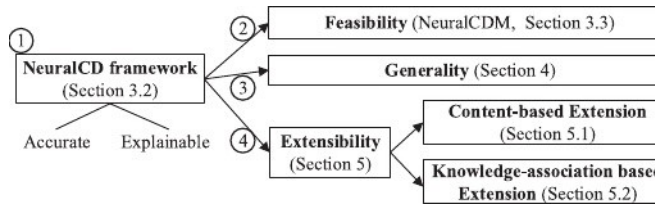


图 3。这项工作的组织。

在基于知识关联的扩展中，我们提出了一种结构扩展 NeuralCD 来解决知识概念覆盖问题。具体而言，提出了一种基于表示的方法来捕获不同知识概念之间的关联，以提高学生对未揭示知识概念的诊断结果的可靠性。这两种类型的扩展可以组合在一起以获得更好的性能。

我们的工作安排见图 3。我们首先在 3.2 节介绍了 NeuralCD 框架。然后在第 3.3 节中，我们用一个已实现的模型 NeuralCDM 来证明 NeuralCD 的可行性。之后，分别在第 4 节和第 5 节介绍了 NeuralCD 的通用性和可扩展性，并从基于内容的可扩展性和基于知识关联的可扩展性两个方面进行了进一步的讨论。

最后，我们在真实世界的数据集上进行了大量的基本和扩展实现实验，结果显示了 NeuralCD 框架在准确性和可解释性方面的有效性。

我们的代码可以在 [https://github.com/bigdata-ustc/neural\\_cognitive\\_diagnostics-neuralcd](https://github.com/bigdata-ustc/neural_cognitive_diagnostics-neuralcd) 上找到

2 相关工作

在本节中，我们从问题难度预测[28]、代码等应用方面简要回顾了相关工作。教育[29]和公式图像转录[30]。如何

认知诊断。现有的关于学生认知的研究中，利用神经网络进行认知诊断是一种非定性的诊断方法，主要来源于教育心理学领域，在参数解释方面表现较差。DINA[9]、[17]和 IRT[10]是其最具典型性的两个固有特征。据我们所知，deep cal works 对学生回答知识追踪(DKT)[31]的结果进行建模，这是使用循环神经网络、学生(uu)和练习(bb)将练习建模为学生学习过程的特征特征之间的交互作用的第一次尝试。具体来说，在 DINA 中，u 后面跟着一些变奏曲[32]，[33]。但是，这些和 b 都是多维的和二元的，其中 b 是知识追踪模型，它更多的是直接从 q 矩阵(人类标记的练习知识-学生状态的变化来预测学生的成绩，边缘相关矩阵)建模。另外两个练习因素，即没有明确地模拟学生的知识猜测和滑动(参数为 g 和 s)的影响，也是考虑到教育的学习过程中的熟练程度。学生的概率为基础。因此，这些模型不适合进行认知诊断，即以 Pðrij¼1juuiP¼nosis 为模型，对运动进行正确回答。很少有神经网络的作品具有高的解释- g1?Hjjð1 ?SjPhij, hij¼k ubik jk。另一方面，在能力上对学生进行认知诊断。为此我们提出了一种神经认知诊断方法(Neu- traits，表征学生的能力和运动难度)。从教育中借鉴特征间相互作用概念的 rcd 框架，以心理模型的方式将其与拟合函数逻辑相结合，如简单版本为 sigmoidðaðu u ?BðP，从数据中学习。NeuralCD 可以实现高精度度-其中 a 是运动判别参数。虽然生动和解释与神经网络。此外，在 IRT 中增加了额外的参数[18]、[19]，潜在框架具有普遍性，可以覆盖许多传统模型，并将特征扩展到多维(MIRT)[11]、[20]，同时易于扩展。

他们的大多数项目响应功能仍然是类似于物流的。知识覆盖问题。知识覆盖范围是

近年来，从数据挖掘的角度研究认知诊断中的一些重要问题。传统的认知模型已经证明了 MF 在认知诊断模型中的可行性。学生与习题对应的是用户与习题的量，习题的知识概念是小矩阵分解(MF)。例如，Toscher 等人的[21]和学生的反应是完整的。知识

改进 SVD (Singular Value Decomposition)方法对分数矩阵进行因子分解，得到学生和练习题的潜在特征向量。Thai-Nghe et al.[22]，[23]应用了一些推荐系统技术，包括在教育背景下的矩阵分解，并将其与传统的回归方法进行了比较。

这些传统模型的交互功能都是手工设计的，是基于各种教育或心理计量学的理论或假设。例如，Reckase 总结了大多数 IRT/MIRT 模型[11]所采用的假设，包括学生之间的独立性、学生和测试中练习的不变性、单调性假设等。在 DINA 模型中，假设一个学生只有在两种情况下才能正确地回答问题：一种是学生已经熟练地掌握了该题所要求的所有技能，没有出错；另一种是学生没有掌握所有技能，但成功地猜出了答案。由于所选择的理论/假设，传统的认知诊断模型可能在某些情况下表现良好。然而，应用范围受到限制，认知过程的过度简化会导致拟合能力有限。在实践中，研究人员需要从各种选择([24])中选择合适的模型，甚至需要为特定的用途设计自己的模型，这是劳动密集型的。虽然专业的理论和假设为认知诊断提供了有价值的建议，但我们需要一种新的诊断模型，它需要较少的专家知识(即自动学习)，并提供准确、可解释、易于理解的诊断结果。

人工神经网络。人工神经网络技术在语音识别[25]、文本分类[26]和图像字幕[27]等领域已经达到了最先进的水平。也有一些教育意义

授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer. 限制适用。



在这些条件下，每个学生的保险是完整的。然而，当习题量和知识概念量大而响应量稀疏时，这是当今智能教育系统中常见的情况，知识覆盖问题就变得不可忽视。传统的模型如 IRT[10]和 MIRT[11]、[20]没有考虑知识概念。在 DINA[9]、[17]、DINO[34]和 NIDA[35]等模型中，知识概念被认为是独立的。当知识覆盖不完全时，这些模型的诊断结果可靠性较低。一些研究考虑了不同知识概念熟练程度之间的关系。例如 AHM[36]考虑了知识概念之间的层次关系。De La Torre et al.[37]提出了 HO-DINA，它考虑了影响学生对每个知识概念熟练程度的低维高阶潜特征。Liu et al.[38]提出了一个 FuzzyCDF 模型，学生对知识概念的熟练程度受其能力参数的影响。然而，据我们所知，在现有的著作中，知识覆盖问题还没有得到明确的研究。

3 神经认知诊断

我们首先正式引入认知诊断任务。然后对 NeuralCD 框架进行了详细的描述。然后，我们利用传统的  $q$ -矩阵设计了一个具体的诊断网络 NeuralCDM，以证明该框架的可行性。在接下来的两节中，我们将从通用性和可扩展性两个方面介绍 NeuralCD 框架的优越性。

3.1 任务简介

假设一个学习系统中有  $N$  个学生、 $M$  个练习和  $K$  个知识概念，这些概念可表示为  $S = \{s_1, s_2, \dots, s_N\}$ ;  $E = \{e_1, e_2, \dots, e_M\}$  和  $K = \{k_1, k_2, \dots, k_K\}$  分别为  $G$ 。每个学生选择一些练习进行练习，反应日志  $R$  表示为三连音  $\delta_s$  集合  $\{e, r\}$  其中  $s \in S$ ;  $e \in E$  和  $r$  是学生在练习  $e$  中得到的分数(转成百分比)。此外，我们还有  $Q$  矩阵(通常由专家标注)  $Q = [Q_{ij}]_{M \times K}$ ，其中  $Q_{ij} = 1$  如果运动  $e_i$  涉及知识概念(简称 KC)  $k_j$  和  $Q_{ij} = 0$ ，否则。

问题定义。根据学生的反应日志  $R$  和  $Q$  矩阵  $Q$ ，我们的认知诊断任务的目标是通过学生表现预测过程来挖掘学生对知识概念的熟练程度。

3.2 神经认知诊断框架

一般来说，认知诊断模型是用来模拟学生的练习回答过程的结果，学生利用自己的认知(如知识、技能)去克服练习中设置的障碍。因此，对于一个认知诊断系统，基本上需要考虑三个要素:学生因素、运动因素以及它们之间的交互作用[39]。在本文中，我们提出了一个通用的 NeuralCD 框架，通过使用多层神经网络建模来解决这些问题，如图 4 所示。具体来说，对于每个响应日志，我们使用 one-hot

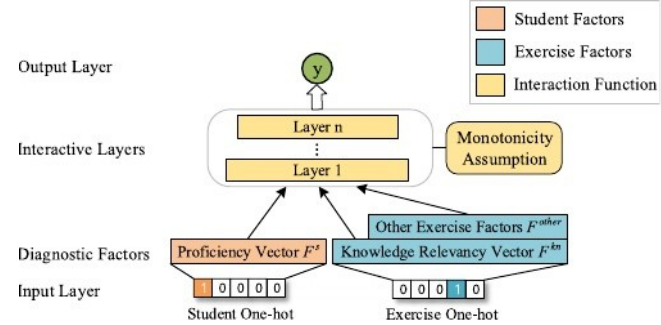


图 4。NeuralCD 框架结构。

将相应的学生与习题向量作为输入，得到学生与习题的诊断因子。交互层学习各因素之间的交互作用，输出正确回答题的概率。训练后，我们得到学生的熟练度向量作为诊断结果。具体介绍如下：

学生因素。学生因素是学生的特点，它会影响学生对练习的反应。由于我们的目标是挖掘学生对知识概念的熟练程度，所以我们没有使用 IRT 和 MIRT[11]中的潜在特征向量，这不足以解释指导学生的自我评估。相反，我们将学生因子设计为类似 DINA 的可解释的向量，但有一个主要的区别，即它们是连续的。具体来说，我们用向量  $F^s$  一个学生的特征，即熟练度向量。 $F$  的每一项  $s$  是连续的  $([0,1])$ ，表示学生对某一知识概念的熟练程度。例如， $F^s = [0.9; 0.2]$  表示对第一个知识概念的掌握程度高，对第二个知识概念的掌握程度低。 $f^s$  是通过参数估计过程得到的。

运动因素。运动因子的设计是为了表征运动的特性。我们把运动因素分为两类。第一个表示练习和知识概念之间的关系，这是我们需要它使  $F$  的每一个条目  $s$  对应于我们诊断目标的特定知识概念。我们称其为知识关联向量，并表示为  $F^{kn}$ 。 $f^{kn}$  和  $F$  的维数相同  $s$ ，其中第  $i$  项表示练习与知识概念  $k$  之间的相关性  $i, k$ 。 $F$  的每一项  $kn$  是非负的。 $f^{kn}$  (例如，由  $q$  矩阵得到)。其他因素属于第二种类型，是可选的。IRT[19]和 DINA[9]中的知识困难、运动困难和辨别能力等因素，在合理的情况下可以加以考虑。

交互功能。交互功能模拟学生因素与运动因素之间的交互作用，得到响应结果(如对或错)。我们使用人工神经网络来获得交互函数，原因如下:首先，神经网络已被证明能够逼近任何连续函数[40]。神经网络具有较强的拟合能力，能够较好地捕捉学生与运动因素之间的关系。其次，利用神经网络，交互功能可以从数据中学习，假设很少(传统模型背后)。这使得 NeuralCD 更加普遍，可以应用于更广泛的领域。第三，该框架具有高度的神经网络可扩展性。例如，

额外的信息，如练习文本，可以集成到神经网络中(我们将在下面的小节中讨论它的可扩展性)。在数学上，我们将 NeuralCD 框架的输出表示为

$$y = \varphi_n(\dots \varphi_1(F^s, F^{kn}, F^{other}, \theta_f)), \tag{1}$$

哪里 $\varphi_i$ 为第*i*个MLP层的映射函数; $f^{other}$ 为F以外的因子 $s$ 和 $F^{kn}$ (如难度);还有 $u_i$ 表示所有交互层的模型参数。

可解释性保证。获得可解释的结果是认知诊断不可或缺少的，诊断结果是评估学生学习状态和进一步提供个性化支持的基础。然而，由于神经网络固有的一些特性，通常对[41]的解释性能较差。为了解决这个问题，我们采取了两个步骤来确保这些因素是可以解释的。第一步是求 $F^s$ ? $f^{kn}$ 在输入层中，以对齐F的每个维度 $s$ 到一个由F的相应维度指定的知识概念 $kn$ 。第二步是利用在某些IRT和MIRT模型[11]中使用的单调性假设，使F中的值 $s$ 单调性假设在几乎所有情况下都是普遍且合理的，因此对NeuralCD框架的通用性影响较小。假设定义如下：

单调性假设。在学生的知识熟练程度的任何维度上，正确回答练习的概率都是单调递增的。

这个假设应该转换成交互函数的一个属性。直觉上，我们假定学生能正确地回答练习*e*。在训练过程中，如果模型输出了错误的预测(即值小于0.5)，优化算法应该提高(或至少不降低)学生的熟练度。每个知识熟练度的增量由F控制 $kn$ (步骤1)。

在介绍了NeuralCD框架的结构之后，我们将展示一些具体的实现。我们首先实现了一个基于NeuralCD的基本模型，其中知识关联向量直接从预先给定的Q-矩阵中得到，以显示NeuralCD的可行性(第3.3节)。然后，我们讨论了NeuralCD的通用性，表明一些传统模型可以被视为它的特殊情况(第4节)。进一步，我们从内容方面(第5.1节)和结构方面(第5.2节)展示了NeuralCD的可扩展性。

3.3 神经认知诊断模型

本文介绍了在NeuralCD框架下的神经认知诊断模型(NeuralCDM)。图5显示了NeuralCDM的结构。

学生因素。在NeuralCDM中，每个学生用一个知识熟练度向量表示。学生因子 $F^s$ 上述为 $h^s$ 这里和 $h^s$ 是通过乘以一个学生的一个热表示向量 $x$ 得到的吗 $s$ 一个可训练的矩阵 $a$ ，即，

$$h^s = \text{sigmoid}(x^s \times A), \tag{2}$$

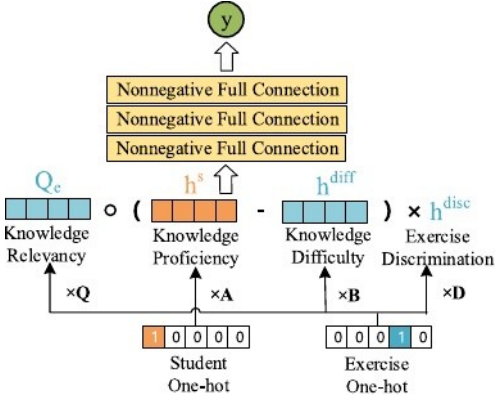


图 5。神经认知诊断模型。

运动因素。对于每项运动，上述运动因子 $F^{kn}$ 是 $Q_e$ 这里，它直接来自于预先给定的 $q$ 矩阵

$$Q_e = x^e \times Q, \tag{3}$$

其中 $Q_e \in \{0,1\}^{1 \times K}$ ， $x^e \in \{0,1\}^{1 \times M}$ 是运动的一个热点代表。为了做出更准确的诊断，我们还采用了另外两个运动因素：知识难度和运动强度 $diff$ 运用辨别能力 $h^{disc}$ 。 $h^{diff} \in \{0,1\}^{1 \times K}$ ，表示该练习所考查的每个知识概念的难度，该难度由IRT中使用的练习难度扩展而来。 $h^{disc} \in \{0,1\}^{1 \times M}$ ，表明该练习能够区分知识掌握程度高的学生和知识掌握程度低的学生。它们可以通过

$$h^{diff} = \text{sigmoid}(x^e \times B), B \in \mathbb{R}^{M \times K} \tag{4}$$

$$h^{disc} = \text{sigmoid}(x^e \times D), D \in \mathbb{R}^{M \times 1}, \tag{5}$$

其中B和D是可训练矩阵。

交互功能。交互层的第一层受到了MIRT模型的启发。我们把它写成

$$x = Q_e \circ (h^s - h^{diff}) \times h^{disc}, \tag{6}$$

在哪里 $\circ$ 是元素明智的产品。以下是两个完整的连接层和一个输出层：

$$f_1 = \phi(W_1 \times x^T + b_1), \tag{7}$$

$$f_2 = \phi(W_2 \times f_1 + b_2), \tag{8}$$

$$y = \phi(W_3 \times f_2 + b_3), \tag{9}$$

其中 $\phi$ 为激活函数。这里我们用的是Sigmoid。

可以使用不同的方法来满足单调性假设。我们采用一个简单的因此，期间总是满足单调性假设

训练。

NeuralCDM的损失函数是交叉熵之间的

其中 $h^s \in \{0,1\}^{1 \times K}$ ; $x^s \in \{0,1\}^{1 \times N}$ ; $a \in \mathbb{R}^{N \times K}$ 。输出 $y$ 和真实标签 $r$   
授权许可使用范围:清华大学。下载于2023年8月05日14:04:04 UTC从IEEE xexplorer。限制适用。

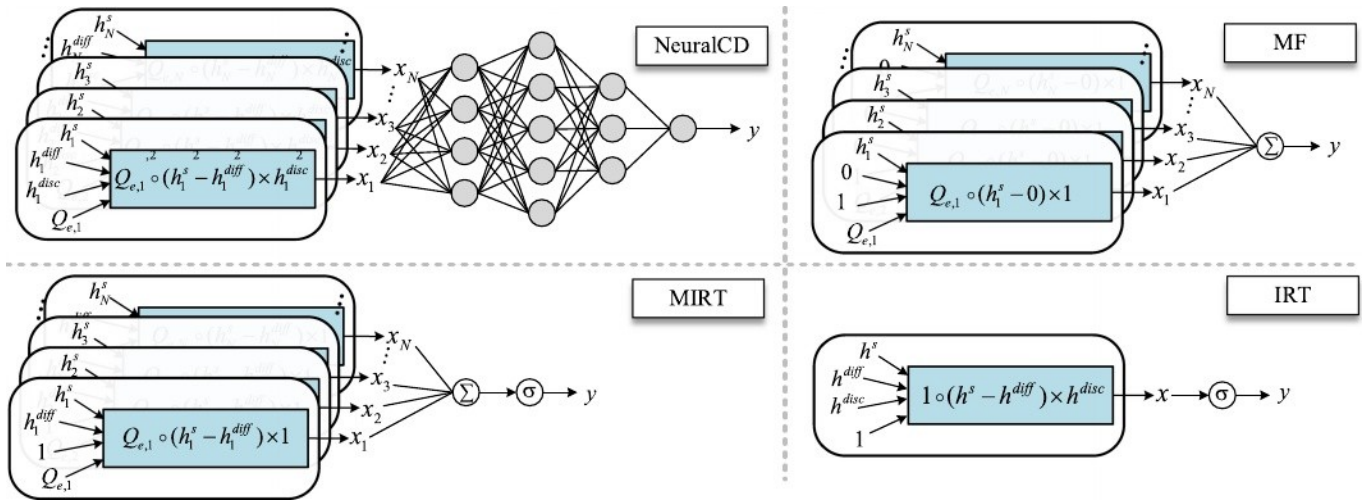


图 6。NeuralCD 与传统模型的关系。

$$loss_{CDM} = - \sum_i (r_i \log y_i + (1 - r_i) \log (1 - y_i)).$$

训练后，h 的值<sup>s</sup>是我们得到的诊断结果，表示学生的知识熟练程度。

4. 神经性疾病的普遍性

在本节中，我们将展示 NeuralCD 是一个可以涵盖许多传统认知诊断模型的总体框架。以式(6)为第一层，我们现在展示了 NeuralCD 与传统模型，包括 MF、IRT 和 MIRT 之间的密切关系。图 6 给出了 NeuralCD 和这些模型的直观比较。

mf[22]。在 MF 中存在学生潜向量和练习潜向量(hh<sup>s</sup> 和 Q<sub>e</sub>)，并取交互作用函数 Q 的基本形式<sub>e</sub>?h<sup>s</sup> 举个例子。应该注意 Q<sub>e</sub>MF 是一个可学习的运动参数，不能表示知识的相关性。通过设置 h<sup>diff</sup>?0 和 h<sup>disc</sup>?1、第一层输出为 x<sub>1</sub>?Q<sub>e</sub>?h<sup>s</sup>。然后，为了像 MF 一样工作 p，所有其余的层需要做的是 sum() x 中每个条目的值，这是很容易实现的。MF 方法中没有采用单调性假设。

太好了。以典型的 IRT y<sub>i</sub>?Sigmoid(h<sub>i</sub>) 的形成为例<sup>s</sup>?h<sup>diff</sup> p?h<sup>disc</sup> 例如:p。首先，集合 Q<sub>e</sub>?1，令 h<sup>s</sup> h<sup>diff</sup> 由于是单维的，第一层的输出为 x<sub>1</sub>?Q<sub>e</sub>?h<sup>s</sup>?h<sup>diff</sup> p?h<sup>disc</sup>。其次，NeuralCD 中的多层神经网络退化为单一的 Sigmoid 激活函数(s)，通过限制 h 可以实现单调性假设<sup>disc</sup>要积极。IRT 的其他变体(如 y<sub>i</sub>?C/(1 + exp(-C(h<sub>i</sub> - b<sub>i</sub>)))，其中 C 是猜测参数)可以实现与一些变化。

我很高兴。一个直接的从 IRT 到 MIRT 的扩展是使用练习和学生的多维特征向量。这里我们以[20]中提出的典型地层为例

$$y = \frac{e^{Q_e \cdot h^s - d_e}}{1 + e^{Q_e \cdot h^s - d_e}},$$

其中 Q<sub>e</sub> 低维参数通常是从响应数据而不是之前给出的

授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。

(10) 知识关联向量。首先，让 h<sup>disc</sup>?1, Eq.(6)给出的第一层输出为 x<sub>1</sub>?Q<sub>e</sub>?h<sup>s</sup>?h<sup>diff</sup> p。第二，使 NeuralCD 中的 p 多层退化为一个求和()，然后是一个 Sigmoid 函数(s)。具体来说，通过设置 W<sub>1</sub> 1 / 4 ??1 ?b<sub>1</sub> 式 (7) 中的 1/40 和 f(x)=1/(1+exp(-x)) 有 f<sub>1</sub> 1/4 Sigmoid(Q<sub>e</sub>?h<sup>s</sup>?h<sup>diff</sup> p?h<sup>disc</sup>)。f<sub>1</sub> 可以在没有更多层的情况下输出。单调性假设不是强制性的，但当 Q 的每一个元素都是单调性时，可以实现单调性假设。被限制为非负的。

5. 神经递质的延展性

在本节中，我们将展示 NeuralCD 是一个易于扩展的开放框架。我们演示了两种类型的扩展，即利用额外的内容和探索知识关联。

5.1 基于内容的扩展

由于手工设计交互功能的局限性，传统的认知诊断模型大多集中在学生 id、习题 id、答题结果(对或错)以及学生答题日志中习题知识概念等数值数据上。然而，这些信息不足以描述学生的认知过程，这是相当复杂的。许多其他的信息，如学生回答习题的时间长短、习题的文本内容等，也与学生的回答有关，这些信息已被证明与一些习题特征(如难度、相关知识概念[28]、[42])高度相关。因此，一个可扩展的框架应该能够聚合这些额外的信息以获得更好的诊断结果。在此，我们选择了练习文本中两种典型的信息类型，即知识概念和文本外相关因子，并说明了它们的应用。

5.1.1 从文本内容中提取知识

第一个演示是从练习内容中提取相关的知识概念。在 (11) NeuralCDM 中，我们使用手工标记的 q 矩阵来表示知识

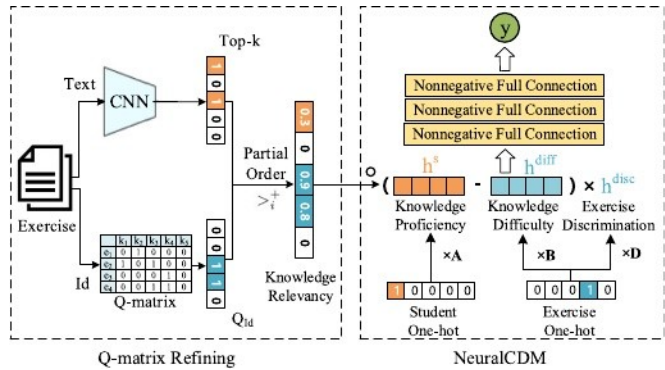


我们定义  $p_{\delta a} > p_{\delta b} Q_{\sim, \delta b}$  具有一对逻辑类功能

$$p(a >^+_i b|\tilde{Q}_i) = \frac{1}{1 + e^{-\lambda(\tilde{Q}_{ia} - \tilde{Q}_{ib})}}. \tag{13}$$

参数是什么?控制标签知识概念与非标签知识概念之间相关性值的辨别。D 上的对数后验分布  $v$  的表达式为

$$\begin{aligned} \ln p(\tilde{Q}|D_V) &= \ln \prod_{(i,a,b) \in D_V} p(a >^+_i b|\tilde{Q}_i)p(\tilde{Q}_i) \\ &= \sum_{i=1}^M \sum_{a=1}^K \sum_{b=1}^K I(a >^+_i b) \ln \frac{1}{1 + e^{-\lambda(\tilde{Q}_{ia} - \tilde{Q}_{ib})}} \\ &\quad + C - \sum_{i=1}^M \sum_{j=1}^K \frac{\tilde{Q}_{ij}^2}{2\sigma^2}, \end{aligned} \tag{14}$$



图七。CNCD-Q: 经过 q 矩阵细化的内容增强 NeuralCD。

每个练习的相关性(传统作品中的常见做法)。但是手工标注的 q 矩阵可能存在缺陷, 因为不可避免的误差和主观偏差[39], [43]。例如, 在 q 矩阵中, 可能只有“方程”被标记为求解方程的练习。但是, 如果我们在文本中发现“”的存在, “分裂”也是必须的。q 矩阵中只标注了目标知识概念而忽略了其他相关知识概念的现象十分普遍。解决这一问题的一种可选策略是利用文本内容, 通过发现练习题中被忽略的知识概念来细化 q 矩阵, 利用神经网络的优势是可行的。我们将此扩展模型表示为 q-矩阵细化内容增强的 NeuralCD (CNCD-Q), 其结构如图 7 所示。

具体来说, 我们首先对一个模型进行预训练, 该模型预测与输入练习相关的知识概念。对于这个任务[44], 可以使用许多适合于文本处理的模型。由于 CNN (convolutional neural network, 卷积神经网络)在文本处理中具有提取局部信息的优势, 它能够从文本中捕获重要的单词(例如与特定知识概念高度相关的单词)。因此, CNN 对于我们的目标来说实际上是足够的。比较或设计更先进的 KC 预测模型超出了这项工作的范围, 我们把它留给未来的研究。我们使用的 CNN 网络以连接在文本中单词的 word2vec 嵌入作为输入, 并将每个预定义的知识概念(在数据中发生的)的相关性输出到练习中(详见 6.2 节)。用人标记的 q 矩阵作为训练的标记。我们定义  $V_i^k = \{v_{ij1}, v_{ij2}, \dots, v_{ijk}\}$  为练习  $e$  的 top-k 知识概念集, CNN 的报道。

然后结合  $V_i^k$  用 q 矩阵。虽然人标 q 矩阵存在缺陷, 但仍然具有较高的置信度。因此, 我们认为用 Q 矩阵标记的知识概念相对于  $i k$  更具有相关性  $j k; 2 v_i^k$  和  $Q_{ij} \frac{1}{4}$  克。为此, 我们采用如下的两两贝叶斯方法。为了方便起见, 我们定义部分顺序  $> p_i$  作为

$$a >^+_i b, \text{ if } Q_{ia} = 1 \text{ and } Q_{ib} = 0 \text{ and } b \in V_i^k, \tag{12}$$

定义偏序关系集为  $D_v 1 / 4; a; B p_j a >_{\delta b} b; I / 4; 2; \dots$  毫克。按照传统的贝叶斯处理方法, 我们假设  $Q_{\sim}$  遵循一个零平均高斯先验, 每个维度的标准差  $s$ 。为了提高 q 矩阵标签的置信度, 授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。

C 是一个可以在优化过程中忽略的常数。在 NeuralCDM 中使用  $Q_{\sim}$  之前, 我们需要将其元素限制在 (0,1) 范围内, 并将未标记或未预测的概念元素设置为 0。因此, Sigmoid  $\delta Q_{\delta b} \sim ?$  NeuralCDM 用 M 代替 Q, 其中  $m_2 f_0; 1$  克  $M^2 K$  是掩模矩阵, 而  $M_{ij} j 2 v = \frac{1}{4} 1_{jk}$  或  $Q_{ij} \frac{1}{4}; m_{ij} \frac{1}{4}$ , 否则。  $Q_{\sim}$  与认知诊断模型一起训练, 因此损失函数为

$$loss = -\ln p(\tilde{Q}|D_V) + loss_{CDM}. \tag{15}$$

5.1.2 文本内容的因子提取

抽取相关知识概念并不是利用习题文本内容的唯一途径。其他一些因素, 如猜, 滑[9], [17]和游戏[45]也被认为是影响学生表现的原因。同样, 其他认知相关的信息也包含在练习文本中。例如, 理解文本内容是解决练习的第一步。有时文本中使用的表达可能会令人困惑, 尽管练习中检查的知识概念可能并不困难。这里有两个例子:

?? ? 10 p 3%:

?E2: Alice 的速度是 1 米/秒。约翰的速度是爱丽丝的两倍。约翰从 3 米的直线跑道起点出发, 并向前移动。10 秒后 John 离起点有多远?

E1 和 E2 测试加法和乘法的掌握。E1 的文本很简单。但是, E2 假设了一个实际的场景, 学生首先需要理解并切换到 E1 这样的表达式。虽然 E1 和 E2 测试的知识概念难度相近, 但正确解题的可能性受到学生对课文内容理解的影响。

传统的认知诊断模型难以实现

聚合这类内容信息, 因为它们的可扩展性有限。在这里, 我们展示了一个考虑额外类型的运动因子, 即文本因子的例子, 以扩展我们的 NeuralCD 框架(标记为 CNCD-F)。CNCD-F 的结构如图 8 所示。虚线框内是文本因子提取过程, 给出练习文本, 我们首先使用 TextCNN[46](这对处理 NLP 文本很有效)来获得嵌入  $e_{2r}$  的文本  $d_0$ 。

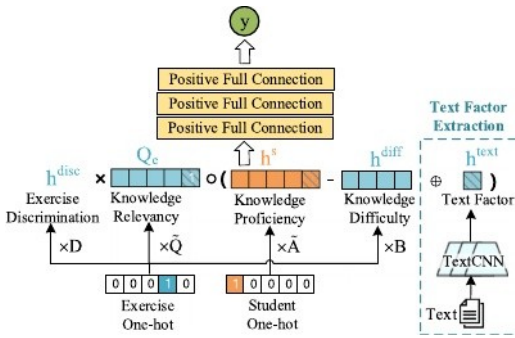


图 8。CNCD-F:含有文本因子的内容增强神经光盘。

然后我们把它转换成文本因子向量  $h^{text} \in \mathbb{R}^{d_1}$  通过

$$h^{text} = W_t \times e + b_t, \tag{16}$$

其中  $W_t \in \mathbb{R}^{d_1 \times D_1}$  和  $b_t \in \mathbb{R}^{d_1}$  是可训练参数，我们设  $d_1 = 1/4$ 。相应地，我们将矩阵  $A$  和  $Q$  推广到  $A \in \mathbb{R}^{N \times K \times d_1}$  和  $Q \in \mathbb{R}^{M \times K \times d_1}$ 。扩展维度表示与文本要素(如阅读理解)相对应的技能。 $d_1 = 1/4$ 。扩展的维度/因子为 -1

在图 8 中，我们使用了带有斜线的正方形。

整个建模过程与 NeuralCDM 类似，除了以下几点：

- 1) Eq.(3)中的  $Q$  和 Eq.(2)中的  $A$  分别用  $\tilde{Q}$  和  $\tilde{A}$  代替。
- 2)将式(6)改为

$$x = Q_e \circ (h^s - (h^{diff} \oplus h^{text})) \times h^{disc}. \tag{17}$$

训练后，第一个  $K$  维的值在  $h^s$  是对 KCs 的诊断熟练程度。

### 5.2 基于知识关联的扩展

在这个小节中，我们展示了 NeuralCD 的结构是可以灵活扩展的，以适应不同的情况。为了解决认知诊断中的知识覆盖问题，我们提出了一种基于知识关联的扩展方法。通常情况下，在一个智能教育系统中，对于一个单一的学科可能会有许多知识概念。但由于学生回答的问题有限(测试或诊断时间较短)，该学生对知识概念的覆盖率通常较低(对真实数据集的分析见图 9 和表 2)。为方便起见，我们称学生的这些未经测试的 kc 为弱 kc，而那些包含的 kc 中超过一半是弱 kc 的题目为弱 kc。在我们以前的模型中，这些弱 kcs 的熟练度值是不可靠的。例如，如果学生从来没有回答过一个与统计学相关的问题，那么学生的知识熟练度向量中相应的维度( $h^s$ )永远不会被训练，因为相关性总是 0。为了使诊断模型更加健壮，有必要对学生的弱 kcs 进行熟练程度近似。

在本研究中，我们通过考虑知识概念之间的关系来解决这个问题。现有的研究

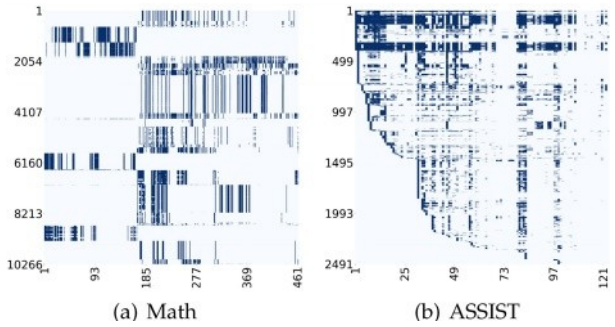


图 9。学生反应日志的知识覆盖率。

揭示了知识概念不是独立的[36]，[47]。知识熟练度与习题的知识难度相互关联。我们计算了学生的熟练程度，关于知识  $k_i$  作为教授  $j, k_j \in \mathbb{F}$  的  $\delta_{s_i, k_j} \text{Rel}^k$ ， $\text{Rel}^k$  表示知识关系，如知识层次[36]、知识概念图[48]。考虑到这些显式知识关系需要专家知识，并且并不总是可用的，我们在这里提供了一个纯粹从响应日志建模隐式知识关系的例子。在缺乏显式知识关系的情况下，通常的做法是计算每对项目之间的成对约束，这需要大量的参数，特别是当约束是针对学生或针对练习的时候。相反，我们采用一种基于表示的方法来隐式建模知识关联，扩展的框架称为 KaNCD。

具体来说，我们不直接学习矩阵  $A$  (式(2))。每名学生( $i$ )及各 KC ( $k_j$ ) 分别用  $d$  维( $d < K$ )潜向量( $l_{si}$  和  $l_{kj}$ )表示。 $A$  中的每个元素都是对应的 student 与 KC 向量运算的结果。这里我们将  $d$  维视为预定义知识概念(受[37]启发)背后的高阶技能。 $l_{kj}$  中每个维度的值表示其对每个高阶技能的偏好，因此  $l_{si}$  被

$$a_i = l_{si} \cdot l_{kj}; \tag{18}$$

然后  $s$  的熟练程度  $i$  关于  $k_j$  被计算为过滤的潜在特征的加权和，Sigmoid 激活(在式(2)中，将熟练程度限制在(0,1))

$$A_{i,j} = W_{a2} \times a_i + b_{a2}, \tag{19}$$

其中  $W_{a2} \in \mathbb{R}^{d_1 \times d_1}$  还有  $b_{a2} \in \mathbb{R}$  是可训练参数。

这样，我们将同样的过程应用到考虑知识关联的知识难度矩阵  $B$  上。每个练习( $e$ )用  $d$  维潜向量( $l_{ei}$ )表示。 $e$  的难度  $i$  在  $KC_j$  计算方式为

$$b_i = l_{ei} \cdot l_{kj}, \tag{20}$$

$$B_{i,j} = W_{b2} \times b_i + b_{b2}, \tag{21}$$

其中  $W_{b2} \in \mathbb{R}^{d_1 \times d_1}$  还有  $b_{b2} \in \mathbb{R}$  是可训练参数。

总的来说，KaNCD 的流程如下。我们计算  $A_{ij} \in \mathbb{R}^{n \times J \times 1}; \dots; K \times P$  和  $B_{ij} \in \mathbb{R}^{m \times J \times 1}; \dots; K \times P$ ，得到  $A$  和  $b$ ，然后用方程式输入训练数据。(2)、(3)、(4)、(5)、(6)、(7)、(8)、(9)对学生进行培训



与式(10)损失函数相同的参数(包括学生潜向量、练习题和 KCs)。训练结束后, 可以用方程式推断学生对 KCs 的熟练程度。(18)、(19)和式(2)。在附录 C 中, 我们将 KaNCD 与 AHM[36]、HO-DINA[37]、FuzzyCDF[38]等现有相关模型进行了比较。

5.3 讨论

我们详细介绍了 NeuralCD 框架, 并展示了它的一些特例。NeuralCD 是一个通用的框架, 可以得到准确和可解释的诊断结果。同时, 该框架具有比传统模型更好的可扩展性如聚合额外信息(如文本内容)和改进框架结构以解决具体问题(如知识覆盖问题)。1)有必要指出, 学生的熟练度向量  $F^s$  和练习的知识关联向量  $F^{kn}$  是 NeuralCD 框架所需的基本要素。在合理的情况下, 可以将运动歧视等其他因素纳入考虑范围。2)第一交互层的形成不受限制, 最好包含  $F^s$  和  $F^{kn}$  以保证  $F$  的每个维度  $s$  对应于特定的知识概念。3)非负全连接只是实现单调性假设的策略之一。更复杂的网络结构可以设计为交互层。例如, 可以使用递归神经网络或记忆网络来捕捉学生学习过程的时间特征。4)在模型输出方面, 我们关注的是客观练习, 在本文中答案正确(1)或错误(0), 因此 NeuralCD 模型的输出是学生正确回答的概率。事实上, NeuralCD 模型也可以处理非二分法反应的练习。例如, 对于具有连续反应标签的练习(例如计分率在范围(0,1)), 模型输出是预测分数;对于可能得分多的练习, 输出层可以改变为输出一个分类向量, 表示预测的类别(即得分)。建模中可以考虑更好的度量方法, 例如练习[11]中的多个独立组件或多个顺序步骤, 我们将其留作以后的研究。

6 个实验

在本节中, 我们从多个方面进行了大量的实验, 以证明我们的 NeuralCD 模型的有效性:(1)针对基线的学生表现预测任务;(2)诊断结果解释的模型分析;(3)知识概念、习题和学生的学习嵌入的可视化。

6.1 数据集描述

我们在实验中使用了两个真实世界的数据集, 即 Math 和 ASSIST。科大讯飞提供的数学数据集来源于目前广泛使用的在线学习系统知学, 其中包含数学练习和

表 1  
数据集摘要

数据集	数学	协助
#学生	102	41
	68	63
#练习	917,4	177
	95	46
#知识概念	14	1
	88	2
		3
#响应日志	864,7	324,5
	22	72
#每个练习的知识概念	1	1.
	.	1
	5	9
	3	
avg#	2	8.
日志	.	0
	2	5
	8	
STD#log	0.	0.3
>	30	16
	5	

高中考试记录。ASSIST (ASSISTments 2009-2010“技能构建者”)是 ASSISTments 在线辅导系统[49]收集的一个开放数据集, 仅提供学生的响应日志和知识概念。<sup>2</sup>我们选择了公共更正版本, 消除了之前工作[50]指出的重复数据问题。表 1 总结了数据集的基本统计信息。

预处理。为了保证每个学生都有足够的数据进行诊断, 我们分别筛选了 Math 和 ASSIST response log 在 30 和 15 以下的学生。因此, 对于 Math 数据集, 我们得到 10268 名学生, 其中包含 497 个知识概念的习题 2,507 个用于诊断网络, 剩余知识概念未出现在日志中的习题用于 CNCD-Q 的 q 矩阵提炼部分;在数据集 ASSIST 上, 我们得到 2493 名学生, 17671 个练习和 123 个知识概念。我们对每个学生的响应日志进行了 80%/20%的训练/测试分割。对于 ASSIST, 我们按照与 Math 相同的方式对响应日志进行了划分, 但由于没有提供练习文本, 因此在该数据集上没有对 CNCD-Q 和 CNCD-F 进行评估。所有模型均采用 5 倍交叉验证进行评估。

静态知识熟练度的验证。由于数据集由考试日志组成, 学生在数学方面的知识熟练程度比较稳定。然而, 在 ASSIST 中, 学生对某一知识概念的熟练程度可能会发生变化, 因为他将不断地被给予该概念的练习, 直到达到一定的标准(例如, 连续回答 3 个相关练习)。为了分析静态模型(如 NeuralCD 和静态传统模型)是否适用于 ASSIST, 我们比较了 Math 和 ASSIST 的两个度量。第一个度量是每个学生为每个知识概念所做的平均日志量

$$AVG_{\#log} = \frac{\sum_i^N \sum_j^K Log(i, j)}{\sum_i^N \sum_j^K I(Log(i, j) > 0)}, \tag{22}$$

$Log(i, j)$  是学生练习的数量  $i$  回答了有关知识概念  $k_j$ 。此外, 另一个指标是分数  $r$  的平均标准偏差  $i_j$  that  $Log(i, j) > 1$  as

$$STD_{\#log>1} = \text{mean}_{s_i \in S} \left( \text{mean}_{k_j \in K_n, Log(i, j) > 1} (std_{i_j}) \right), \tag{23}$$

2. <https://sites.google.com/site/assistmentsdata/home/assist-1>. <https://www.zhixue.com> -2009-2010-data/skill-builder-data-2009-2010  
授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。

哪里性病<sub>ij</sub> 这个学生的分数的标准差是多少<sub>i</sub> 得到了与知识概念<sub>k</sub> 相关的习题<sub>j</sub>。如表 1 所示，尽管 ASSIST 的 AVG 要大得多<sub>#log</sub> 数学，他们的性病<sub>#log</sub> > 1 很接近。因此，我们可以合理地假设学生在 ASSIST 中的知识状态也是稳定的，我们的静态 NeuralCD 模型和基线都适用于这两个数据集。在模型解释课上会有更多的讨论。

知识覆盖可视化。为了说明每个学生全响应日志的知识覆盖率(之前将数据分为训练集和测试集)，我们绘制了两个数据集的热图。在图 9 中，横轴和纵轴分别表示 KC ID 和学生 ID。蓝色表示对应的学生有相应 KC 相关的响应日志(即回答了相关的练习)，否则颜色为白色。我们可以观察到，在两个数据集上，学生响应日志的知识覆盖率都很低。这印证了我们之前提出的知识概念覆盖问题是一种普遍现象，需要注意的说法。

表 2  
弱题在测试集中所占的比例

	随机分裂		弱覆盖率分割	
	数学	协助	数学	协助
弱响应比例	0.315	0.040	0.8876	0.857

6.3 学生表现预测

认知诊断模型的性能难以评估，因为我们无法获得学生真正的知识熟练程度。在大多数的作品中，诊断结果通常是通过预测学生的表现来获得的，因此这些预测任务的表现可以从一个方面[38]来间接评价模型。为了充分评估模型，我们采用了两种方法来分割数据集。第一种是随机分割，随机抽取每个学生的响应日志的 80%作为训练集，这是学生成绩预测任务中常见的做法。另一个是覆盖范围很窄

设计的目的是为了在 KC 覆盖学生列车时更好地比较模型性能

数据很低。因此，我们设计了一种分割算法。(7)、(8)、

(9)) 分别为 512、256、1,Sigmoid 为

保持训练 / 测试比率 (80%/20%) 不变。详细信息用作所有层的激活函数。我们设置

超参数  $\gamma/0:1$  (Eq.(13)) 和  $s/41$  (Eq. (14)) 在 top-k 知识概念的选择中，我们使用了金属材料。我们设计这个算法是为了使结果表明，预测网络的查全率可达 0.85。

在我们的实验中， $k = 1 / 20$ 。我们初始化了测试集中弱响应比例的差异

对数据大小的差异进行训练。比例参数与 Xavier 初始化 [51]，其中填充

从 N 80 中采样随机值的权重 ;std2p, of responses to weak-questions in test sets are showed in qffffffffffffffffffff

结果被提交给。对于单调性假设，考虑到我们在数据中使用的所有练习都是实现的，并不局限于特定的方法。在客对

在 CNCD-Q 中，CNN 在表 3 和表 4 中分别包含 3 个卷积层，其中 '' 之后的误差条后面是一个全连接输出层。MaxPo

卷积层分别为 400、200、100 和内核大小。从表 3 中，我们将下列值分别设置为 3、4、5。我们采用了 ReLu 活化

在 CNCD-F 中，TextCNN 架构基本上是 CNCD-F) 在 NeuralCDM 之上，证明了额外的信息与 [46] 相同。我们设置的 15

为了评估 NeuralCD 模型的表现，CNCD-F 显示文本因素确实发挥了重要作用，我们将其与之前的方法，即 DINA，不

表 3  
随机分裂法预测学生成绩的实验结果

模型	数学(随机分裂)						协助(随机分裂)					
	准确性		均方根误差		auc		准确性		均方根误差		auc	
Din a irt mirt PM F	0.5	.0	0.4	.0	0.6	.0	0.6	.0	0.4	.0	0.6	.0
	93	0	87	0	86	0	50	0	67	0	76	0
	0.7	1	0.3	1	0.7	1	0.6	1	0.4	1	0.6	2
	82	.0	87	.0	95	.0	74	.0	64	.0	85	.0
	0.7	0	0.3	0	0.8	0	0.7	0	0.4	0	0.7	0
	93	2	78	1	13	1	01	2	61	2	19	1
	0.7	.0	0.4	.0	0.7	.0	0.6	.0	0.4	.0	0.7	.0
	63	0	07	0	92	0	61	0	76	0	32	0
		1		2		2		2		1		1
		.0		.0		.0		.0		.0		.0
		0		0		0		0		0		0
		1		1		2		2		1		1
神经清洁 cdm	0.792	.002	0.378	.001	0.820	.001	0.719	.008	0.439	.002	0.749	.001
CNCD-Q	0.805	.001	0.368	.002	0.836	.001	0.732	.001	0.424	.001	0.767	.001
CNCD-F	0.802	.001	0.370	.002	0.835	.002						

表 4  
基于弱覆盖分割的学生成绩预测实验结果

模型	数学(弱覆盖分割)						ASSIST(弱覆盖分割)					
	准确性		均方根误差		auc		准确性		均方根误差		auc	
Din a irt mirt PM F	0.2	.	0.5	.0	0.5	.0	0.4	.	0.4	.0	0.5	.0
	23	0	02	0	60	0	71	0	90	0	88	0
	0.6	0	0.4	2	0.6	1	0.6	0	0.4	1	0.6	2
	24	1	67	.0	38	.0	57	1	64	.0	33	.0
	0.6	.	0.5	0	0.5	0	0.6		0.5	0	0.6	0
	20	0	83	1	72	2	37	.	05	1	12	2
	0.5	0	0.5	.0	0.6	.0	0.6	0	0.4	.0	0.7	.0
	96	1	85	0	25	0	25	0	78	0	30	0
		.		2		1		1		2		1
		0				.0						.0
		0				0		.				0
		1				1		0				3
								0				
								1				
神经清洁 cdm	0.735	.002	0.432	.002	0.649	.001	0.710	.003	0.455	.001	0.633	.002
CNCD-Q	0.736	.001	0.430	.001	0.691	.001	0.720	.001	0.435	.001	0.732	.002
CNCD-F	0.748	.001	0.419	.001	0.732	.001						
	0.741	.001	0.419	.001	0.732	.001						

弱知识覆盖场景。从表 4 可以看出:首先,与正常场景相比,模型性能显著下降,表明 KCs 的低覆盖问题对诊断结果有相当大的负面影响。其次,NeuralCD 模型仍然比基线表现得更好,这证明了我们的 NeuralCD 框架的优越性。第三,扩展的 NeuralCD 模型在 NeuralCDM 上的改进比在普通场景下的改进更明显。也就是说,模型性能的下降



小于 NeuralCDM，这证明了扩展方法增加了 NeuralCD 对低 KC 覆盖问题的容错性。CNCD-Q 和 CNCD-F 的性能优于 KaNCD，表明额外的信息(如练习文本)对认知诊断的积极作用大于仅仅改善模型结构。

6.4 诊断结果的可解释性

学生表现预测任务不足以评价认知诊断模型，因为可解释性是认知诊断结果的重要组成部分。具体来说，我们采用同意度(Degree of Agreement, DOA)[54]作为被诊断的学生状态(hh<sup>s</sup>)。这个指标是基于直觉的，如果学生 a 比学生 b 对知识概念 k 掌握得更好，那么学生 a 比学生 b 更有可能正确回答与 k 相关的练习。对于知识概念 k, DOA<sub>ak</sub> 表示为:

3.这个公式修正了[16]中的错误。

$$DOA(k) = \frac{1}{Z_1} \sum_{a=1}^N \sum_{b=1}^N I(F_{ak}^s > F_{bk}^s) \frac{\sum_{j=1}^M I(Q_{jk} = 1) \wedge J(j, a, b) \wedge I(r_{aj} > r_{bj})}{Z_0}, \tag{24}$$

$$Z_0 = \sum_{j=1}^M I(Q_{jk} = 1) \wedge J(j, a, b) \wedge I(r_{aj} \neq r_{bj}), \tag{25}$$

$$Z_1 = \sum_{a=1}^N \sum_{b=1}^N I(F_{ak}^s > F_{bk}^s) I(Z_0 > 0), \tag{26}$$

其中  $F_{ak}^s$  为学生 a 对知识概念 k 的熟练程度。如果陈述为真，则  $I(\text{Statement}) = 1$ ，否则为 0。如果学生 a 和 b 都做了运动 j 且  $r_{aj} > r_{bj}$ ，则  $J(j, a, b) = 1$ ，否则为 0。应该注意的是，如果  $Z_0 = 0$ ，相应的  $\delta_{a;b;k}$  三联体排除在 DOA 计算之外。我们将所有的知识概念平均 DOA<sub>ak</sub> 来评估诊断结果的质量(即模型获得的知识熟练度)。值得注意的是，虽然我们在式(24)中定义的 DOA 忽略了包含多个 KCs 的练习时的协同作用，但它确实一定程度上反映了一种可解释的认知现象。

在传统模型中，我们只比较 DINA，因为 IRT、MIRT 和 PMF 的潜在特征和知识概念之间没有明确的对应关系。此外，我们还对两个简化的 NeuralCDM 模型进行了实验。在第一个约简模型(记为 NeuralCDM\_Q)中，知识关联向量是在无监督训练中估计的，而不是从  $\eta$  矩阵中得到的。而在另一个简化模型(记作 NeuralCDM\_M)中，单调性

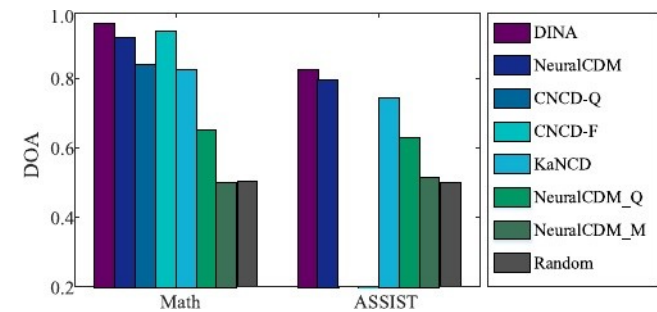


图 10。模型在两个数据集上的 DOA 结果。

通过消除对全连接层的非负限制，消除了假设。利用这两个简化模型分别论证了精细估计知识相关度向量和单调性假设的重要性。此外，我们还进行了一个额外的实验，随机估计学生的知识熟练程度，并计算 DOA 进行比较。

图 10 给出了实验结果，从中我们可以看到几个结论。首先，DINA 模型在两个数据集上都获得了最高的 DOAs，这确实是一个高度可解释的认知诊断模型。这是因为 DOA 测量的认知诊断结果的解释类型与 DINA 背后的直觉高度一致。但是，考虑到学生成绩预测任务的结果 DINA 的认知诊断结果仍然不适合作为诊断报告和进一步的学习辅助。其次，NeuralCD 模型的 DOAs 顺序为 CNCD-F >NeuralCDM >CNCD-Q >KaNCd。虽然比 DINA 低，但其 DOAs 仍显著高于 Random，说明其可解释性强。CNCD-Q 的 DOA 低于 NeuralCDM 的原因是，在 CNCD-Q 中，每个运动的相关 KCs 更多，而 KCs 的增效作用不是用 DOA 来衡量的。KaNCd 的低 DOA 是知识关联建模的结果。CNCD-F 的 DOA 甚至高于 NeuralCDM，说明文本因子是认知模型的良好补充，可在在线补充材料中找到。例如，当学生 A 的知识熟练程度比学生 B 高时，他在某些练习上的分数比学生 B 低，

原因可能是 A 误解了这些练习的文本。由于练习文本的影响被转移到文本因子上，因此 CNCD-F 得到了较高的 DOA。第三，将 NeuralCDM\_Q 和 NeuralCDM\_M 与 NeuralCDM 进行比较，发现其 DOA 值明显下降，说明 q 矩阵信息和单调性假设对获得可解释的诊断结果(知识熟练度向量)都很重要。此外，NeuralCDM 和 KaNCd 在数学上的表现要比 ASSIST 好得多。这主要是由于日志中的矛盾，即学生可能会回答一些包含知识概念 k 的习题 j 正确，而其他包含 k\_j 错(原因可能是知识熟练程度的变化，或者练习中包含的其他知识概念)。如表 1 所示，ASSIST 具有更大的 AVG\_{#log} 性病发病率略高\_{#log > 1} 而不是 Math 数据集，这使得日志中的矛盾更多。更长的原木，更多的矛盾将减少 DOA。

6.5 基于内容扩展分析

练习文本内容为认知诊断提供了有用的补充信息，可在在线补充材料中找到。在表 5 中，我们给出了一个来自 Math 数据集的示例，演示了 CNCD-Q 如何利用文本内容细化 q 矩阵，从而提高诊断性能。从表中可以看出，KC 只有一个，即用 q 矩阵标注的“Number and formula”，这对于描述本次练习所测试的知识概念是不准确的。这种不准确可能是多种原因造成的，比如专家关注的是一项练习的主要知识概念，或者缺乏对知识概念的系统组织。从文本内容中，CNCD-Q 预测与练习相关的 KCs，如“空间中直线与平面之间的位置关系”。对 Q-矩阵标记的 KCs 和预测的 bv CNCD-Q 的相关性进行了区分(分别为 0.87 和 0.45 0.01)。由于我们确定了预测的 KCs 总数(即 20 个)，并受 CNCD-Q 中知识预测组件性能的限制，一些预测的 KCs 可能与演习无关(没有下划线的 KCs)。目前，CNCD-Q 无法区分预测的 KCs 的相关性，这需要进一步改进。

表 5  
q -矩阵求精的一个例子

练习文本内容 设 m;N 是两条不同的直线， a;b;G 是三个不同的平面。 \_\_\_\_ (1) 如果 m k n 和 nparallel a

q -矩阵标记的 KCs(相关性 :0.87) 数字和公式”

CNCD-Q 预测的 KCs(相关性 :0.45) “空间中直线与平面的位置关系”，“位置关系” - 《空间中线的位置关系》、《基本属性与空间关系》、《空间空间关系》 “平面的应用”、“平面间垂直度的判断”、“平面的判断” “线与平面的垂直度”、“线与平面平行度的判断” 《平面之间垂直度的性质》、《垂直度的性质》 《平行线和面的性质》、-《线与面之间的夹角》 《斜线形成的角》、《斜线》、《集》、《简单多面体》、 《边面积、表面积与体积》、《点、线、面之间的距离》、《全量词与存在量词》、 《必要、充分、充分条件判断》

图 11。知识概念嵌入的 t-SNE 可视化。

基于内容的扩展带来的另一个效果是对知识覆盖问题有更好的容忍度。从表 4 的结果可以看出，CNCD-Q 和 CNCD-F 虽然不是专门针对弱知识覆盖场景设计的，但在弱知识覆盖场景下仍然表现良好，有时甚至优于 KaNCD。这可能是因为 CNCD-Q 中，预测的知识概念显著增加了学生响应日志的知识覆盖率。CNCD-F 的扩展尺寸为  $h^s$  对应于文本因子  $h^{text}$  是衡量学生能力的重要指标，无论知识覆盖率问题如何，都影响着学生在所有练习中的表现。综上所述，当有练习文本内容时，利用内容信息是克服知识覆盖问题的较好方法。

6.6 在 KaNCD 中嵌入可视化

利用在 ASSIST 上训练好的 KaNCD 模型，将知识概念的嵌入向量可视化<sup>h)</sup>，通过 t-SNE[56]将它们投射到 2d 点。知识概念嵌入的可视化结果如图 11 所示。我们根据知识概念的位置将它们分成 7 个簇，并用不同的颜色进行区分。这些聚类揭示了一些合理的结果。例如，第一个集群中的知识概念基本上是关于基础代数的。一些相关的知识概念是相近的(如 88 面积矩形和 90 面积三角形)。这些知识关联通过响应日志的嵌入来隐式捕获，这有助于改善弱 kcs 的诊断结果。

讨论。在 KaNCD 中，我们也用向量嵌入来表示每个学生和练习，这也应该捕获重要的信息。嵌入物之间的分布或距离可能反映出一些信息。因此，我们尝试将学生嵌入可视化<sup>s)</sup>和练习嵌入<sup>l)</sup>，使用 t-SNE 方法。为了保证足够的训练数据，我们只可视化嵌入超过 15 个回答的学生和超过 10 个回答的练习。最后我们得到了图 12a 和图 12b，在图中我们可以观察到点遵循特定的模式，而不是随机分散。虽然很难找出它们为什么会出现这种模式的原因，但我们确实发现了一个有趣的现象。如果我们用学生/练习点的正确率(即学生所有回答的正确率和练习中所有回答的正确率)来给这些点上色，我们

可以发现，在图 12a 中，正确的速率从左下到右上逐渐增加，在图 12b 中，正确速率从内到外逐渐增加。这表明低维学生/运动嵌入已经捕获了关于正确率的信息。可能还有一些其他的信息(例如，与知识概念的关系)，我们把这些留给未来的研究。

6.7 案例研究

在图 13 中，我们展示了一个由 NeuralCDM 和 KaNCD 在公共数据集 ASSIST 上给出的学生诊断结果的例子。图 13a 中的表格显示了学生回答的三个练习的 q 矩阵和相应的回答结果。q 矩阵表后面的雷达图是学生对应知识概念的诊断结果。如雷达图所示，NeuralCD 模型可以提供可解释的诊断报告，显示学生对不同知识概念的熟练程度然后，我们比较了学生的知识难度和熟练程度。如图 13a 和 13b 所示，柱状图表示学生对每个相关知识概念的熟练程度，用不同颜色和标记的线表示相关知识概念的知识难度。从两个子图中我们可以看出，当学生答对时，被诊断出的知识熟练程度往往高于知识困难程度。例如，在图 13a 中，练习 3 要求掌握“排序分数”，难度为 0.35。学生回答正确，诊断熟练度为 0.6。知识难度( $hh^{diff}$ )和知识熟练度( $hh^s$ )在我们的模型中是可以解释的。此外，我们在图 14 中绘制了所有由 NeuralCDM 和 KaNCD 诊断的熟练程度。知识概念的运用，使前 45 个概念出现在

图 12。学生和练习嵌入的 t-SNE 可视化。  
授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。



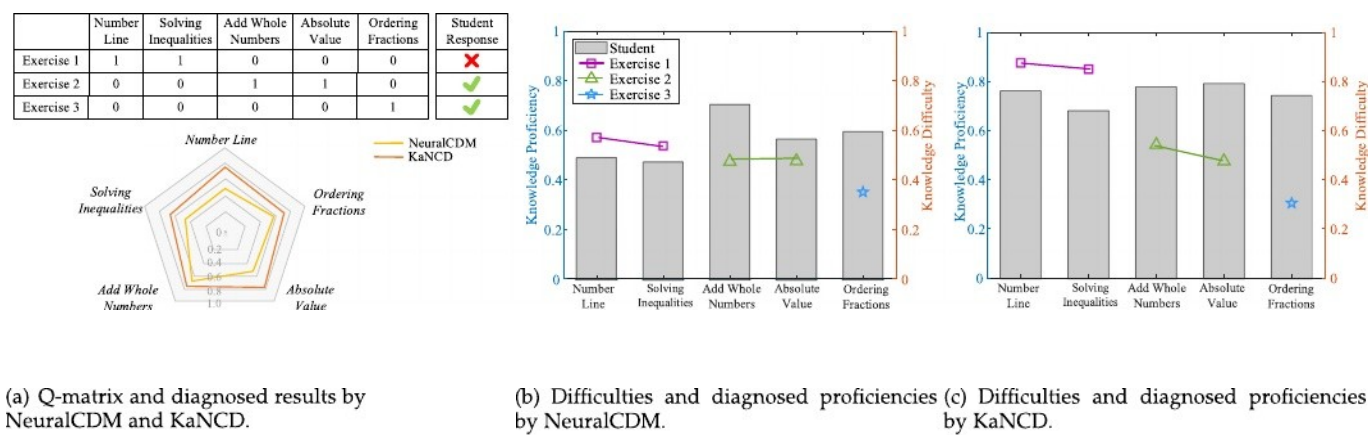


图 13。两名学生在 ASSIST 上的诊断结果及其与知识困难的关系。

学生的训练数据和后 78 个知识概念为弱 kc。我们可以观察到，虽然 NeuralCDM 在前 45 个概念上存在差异，但在弱 kcs 上的熟练度接近 0.5，保持了初始值，从未接受过训练。另一方面，KaNCD 提供的熟练程度在所有知识概念上都有合理的变化，其平均值(0.65)更接近学生的整体正确率(0.75)，这再次证实了 6.6 中的观察，即低维度学生嵌入已经捕获了学生正确率的信息。

进一步观察图 14，我们可以发现模型提供了不同的诊断结果。这引起了一个问题:我们应该相信哪个结果，或者我们应该如何适当地使用诊断结果?由于我们的 NeuralCD 框架和实现的模型是数据驱动的，不同训练过的模型(例如，具有不同的数据或超参数)诊断的熟练程度不能严格保证具有可靠性。诊断的熟练程度的解释和使用应与估计的运动属性(如难度)一起，因为它们是在相同的参数规模。我们将对不同训练模型的熟练程度进行比较，并对其可信度进行验证，以供未来的探索。

### 7 结论与讨论

本文提出了一种神经认知诊断框架——NeuralCD 框架，用于学生的认知诊断。具体来说，我们首先在框架中讨论了基础学生因素和运动因素，并在基于神经网络的框架中放置了单调性假设，以确保诊断结果的准确性和可解释性。然后我们实现特定的模型

在 NeuralCD(即 NeuralCDM 和三个扩展模型)下，通过在真实数据集上的大量实验，我们证明了该框架的可行性。初步表明，神经网络具有较好的认知诊断能力，具有比传统模型更好的应用潜力。我们提出的 NeuralCD 框架不仅对认知诊断具有准确和可解释性，而且具有良好的通用性和可扩展性。具体来说，我们证明了一些传统的模型(如 MIRT)可以被视为 NeuralCD 的特例，而 NeuralCD 可以更灵活地扩展，从而更好地模拟认知过程。例如，可以聚合更多类型的响应数据(如练习文本内容)，可以调整 NeuralCD 的结构以应对不同的情况(如知识覆盖问题)。

认知诊断作为智能教育的基础，为许多自适应学习应用提供支持，如学习反馈、计算机自适应测试[8]、[57]、资源推荐[58]、学习路径规划[59]、[60]。除了智能教育外，认知诊断也广泛应用于需要评估用户能力或心理状态等潜在特征的领域。临床评估是认知诊断[2]的典型应用，包括测量心理障碍[34]、患者报告结果[3]等。在游戏领域中，一个常见的需求是预测玩家的对局和偏好[1]，这需要评估玩家的能力以及合作和竞争能力[61]。在职业领域，An et al.[62]试图评估辩护律师的熟练程度。综上所述，认知诊断是许多领域的基础性工作，也是广泛应用的重要基础。鉴于神经网络具有很高的灵活性和潜力，我们希望这项工作能够对认知诊断在不同领域的进一步研究起到推动作用。

参考文献

S. Chen 和 T. Joachims, “在上下文中预测比赛和偏好”, Proc. 22 ACM SIGKDD Int. Conf. Knowl. 发现. 数据挖掘, 2016, pp. 775-784.

bb0 M. L. Thomas, “项目反应理论在临床评估中的价值:综述”, 《评估》, 第 18 卷, 第 1 期. 3, pp. 291-307, 2011.

J. C. Cappelleri, J. J. Lundy, 和 R. D. Hays, “经典测试理论和项目反应理论的概述, 用于项目的定量评估发展患者报告的结果 meas

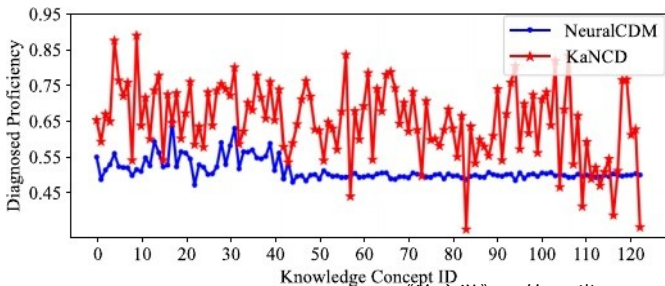


图 14。知识概念的诊断结果。ures, “clinin. 《治疗学》, 第 36 卷. 5, pp. 648-662, 2014.

授权许可使用范围:清华大学。下载于 2023 年 8 月 05 日 14:04:04 UTC 从 IEEE xexplorer。限制适用。

sci. 情报。 , 2019, pp. 782-790.

[4] A. Anderson, D. Huttenlocher, J. Kleinberg , 和 J. Leskovec , “参与大规模的在线课程” , Proc. 23 Int。会议。万维网, 2014, 第 687-698 页。

[5] H. Burns, C. A. luchardt, J. W. Parlett, 和 C. L. Redfield , 《智能教学系统:设计的演变》 , Hove, 英国:心理学出版社, 2014。

b0 R. Wu 等人, “认知模型在预测考生表现中的作用” , 刊于第 24 期。联合会议。情报。 , 2015, 第 1017-1024 页。

G. D. Kuh, J. Kinzie, J. A. Buckley, B. K. Bridges , 和 J. C. Hayek , 拼凑学生成功之谜:研究、主张和建议:ASHE 高等教育报告, 第 116 卷。Hoboken, 新泽西, 美国:Wiley, 2011。

[8] H. Bi 等人, “质量满足多样性:计算机化自适应测试的模型无关框架” , 《IEEE 集成》。数据挖掘, 2020, 第 42-51 页。

bb0 张德华 , “DINA 模型与参数估计:一种教学方法” , 《中国科学院学报》 (自然科学版)。行为。中央集权主义者。 , 第 34 卷 , no. 1, pp. 115-130, 2009。

[10] S. E. Embretson 和 S. P. Reise , 项目反应理论 , Hove, 英国:心理学出版社, 2013。

[11] M. D. Reckase , “多维项目反应理论模型” , 《多维项目反应理论》 , 柏林, 德国:施普林格 , 2009 , 第 79-112 页。

Y. Koren, R. Bell, 和 C. Volinsky , “推荐系统的矩阵分解技术” , 计算机 , no. 8, 30-37 页 , 2000。

Q. Liu , “面向新一代认知诊断” , 刊于 Proc. 30 Int。联合会议。情报。 , 2021, 第 4961-4964 页。

刘新宇, 张敏, 王伟, 刘新宇, “基于图像表示的神经语言模型快速解码” , 中国科学(d 辑:信息科学)系统。 , 2018, pp. 6308-6319。

K. Song, M. Ji, S. Park, and i - c。在 Proc. AAAI Conf. Artif 中, “层次上下文使循环神经网络能够进行推荐”。情报。 , 第 33 卷, 第 1 期。1, 2019, pp. 4983-4991。

[16] F. Wang 等人, “智能教育系统的神经认知诊断” , 《中国科学(d 辑)》。情报。 , 2020, 第 6153 - 6161 页。

bb0 M. von Davier , “作为约束一般诊断模型的 DINA 模型:模型等效性的两种变体” Brit。数学。中央集权主义者。精神病。 , 第 67 卷 , no. 1, 第 49-71 页, 2014。

G. H. Fischer , “rasch 模型的派生” , 摘自《rasch 模型》。柏林, 德国:施普林格, 1995, 第 15-38 页。

刘文杰, 项目反应理论在实际考试中的应用, 北京大学学报, 2012。

R. J. 亚当斯, M. 威尔逊和 w. c。 “多维随机系数多项式 logit 模型” , 清华大学学报(自然科学版)。精神病。米斯。 , 第 21 卷, 第 1 期。1997 年, 第 1 - 23 页。

[21] A. Toscher 和 M. Jahrer , “协同过滤应用于教育数据挖掘” , 《Proc. 4th Int.》。Conf. Educ. 数据挖掘 , 2010, pp. 129-138。

bb0 N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe , 和 L. Schmidt-Thieme , “预测学生表现的推荐系统” 科学。 , vol. 1, no. 1。2 , 页 2811-2819 , 2019。

[23] N. Thai-Nghe 和 L. Schmidt-Thieme , “智能辅导系统中学生建模的多关系分解模型” , 发表于 IEEE 第 7 期。Conf. Knowl。系统。英 , 2015, 第 61-66 页。

p - w。李海华, “认知诊断模型与 q 矩阵选择的拟合指标表现” , 应用研究。精神病。米斯。第 40 卷, 没有。6, pp. 405-417, 2016。

[25] W. Chan, N. Jaitly, Q. Le , 和 O. Vinyals , “听, 参加和拼写:一个用于大词汇量会话语音识别的神经网络” , Proc. IEEE Int. Conf. 声学。 , 语音信号处理。 , 2016, 第 4960-4964 页。

张新宇, 赵建军, “基于字符级卷积网络的文本分类” , 《中文信息学报》 , 第 2 期。系统。 , 2015, pp. 649-657。

Wang W. Wang, Chen Z. , and H. Hu , “图像字幕的分层注意网络” , 《中国科学:信息科学》 , 第 1 期, 第 5 - 7 页。情报。 , 2019, pp. 8957-8964。

[28] Z. Huang 等人, “标准测试中阅读问题的问题难度预测” , 发表在第 31 期 AAAI Conf. Artif. 情报。 , 2017, 第 1352-1359 页。

[29] M. Wu, M. Mosse, N. Goodman, 和 C. Piech , “零机会学习的代码教育:基于深度学习推理的规则抽取” , 刊于第 33 期 AAAI Conf. Artif.

Y. Yin 等人, “用聚束机制从结构图像转录内容”, Proc. 24 ACM SIGKDD Int. Conf. Knowl. 发现. 数据挖掘, 2018, pp. 2643-2652。

bb0 C. Piech 等人, “深度知识追踪”, Proc. advancyninf. Process. 系统. , 2015, pp. 505-513。

bb0 Chen P. , Lu Y. , V. W. Zheng, and Y. Pian, “先决条件驱动的深度知识跟踪”, 《自动化学报》第 1 期. 数据挖掘, 2018, 第 39-48 页。

[33] Q. Liu 等人, “EKT: 学生表现预测的运动意识知识追踪”, IEEE 译. 诺尔. 数据工程, 第 33 卷, 第 1 期. 1, 100-115 页, 2021 年 1 月。

J. L. Templin 和 R. A. Henson, “使用认知诊断模型测量心理障碍”, 《心理学》. 方法, 第 11 卷, 第 1 期. 2006 年 3 月, 艺术. 没有. 287.

bb0 E. Maris, “多分类潜在类模型的估计”, 《心理计量学》, 第 64 卷, 第 1 期. 2, 187-212 页, 1999 年。

J. P. Leighton, M. J. Gierl , 和 S. M. Hunka , “认知评估的属性层次方法:tatsuoka 规则空间方法的变异”J. Educ. 米斯. , 第 41 卷, no. 3, pp 205-237, 2004。

J. d. La Torre 和 J. A. Douglas, “认知诊断的高阶潜在特质模型”, 《心理计量学》, 第 69 卷, 第 1 期. 3, pp. 333-353, 2004。

[38] Q. Liu 等人, “模糊认知诊断在考生表现建模中的应用”, 美国计算机学会汇编. 情报. 系统. 技术. , 第 9 卷, 第 1 期. 4、2018 年, 艺术. 没有. 48.

[39] L. V. DiBello, L. a. Roussos 和 W. Stout, “31a 认知诊断评估的回顾和心理计量模型的总结”, 《统计手册》. , 第 26 卷, 第 979-1030 页, 2006。

K. Hornik, M. Stinchcombe 和 H. White, “多层前馈网络是通用逼近器”, 神经网络. , vol. 2, no. 2. 第 5 页, 359-366 页, 1989。

W. Samek, A. Binder, G. Montavon, S. Lapuschkin 和 k. r. McEuller, “评估深度神经网络学习成果的可视化”, IEEE 译. 神经网络. 学习. 系统. , 第 28 卷, 第 1 期. 11, 第 2660-2673 页, 2017 年 11 月。

[42] Y. Su 等人, “学生表现预测的练习增强顺序建模”, 在 Proc. 32 AAAI Conf. Artif. 情报. , 2018, 第 2435-2443 页。

刘建军, 徐国栋, “数据驱动的 q -矩阵学习”, 应用数学学报. 精神病. 米斯. 第 36 卷, 没有. 7, pp. 548-564, 2012。

W. Huang 等人, “分层多标签文本分类:一种基于注意的循环网络方法”, Proc. 28 ACM Int.。Conf. Inf. Knowl。管理. , 2019, 第 1051-1060 页。

[45] R. Wu, G. Xu, E. Chen, Q. Liu, W. Ng, “知识还是游戏?”基于多次尝试反应的认知模型, ”Proc. 26 Int.。Conf. World Wide Web Companion, 2017, 页 321-329。

“用于句子分类的卷积神经网络”, Proc. Conf. Empir. 方法自然朗. 过程. , 2014, 第 1746-1751 页。

[47] J. Yu 等人, “moocbe: mooc 中 NLP 应用的大规模数据存储库”, 发表于第 58 届会议. 会见协会. 电脑. 语言学, 2020, pp. 3135-3142。

[48] X. Huang 等人, “从多源数据构建多关系的教育概念地图”, 《中国教育科学与工程学报》. 数据挖掘, 2019, pp. 1108-1113。

[49] M. Feng, N. Heffernan , 和 K. Koedinger, “用在线系统解决评估挑战, 在评估过程中进行指导, ”用户模型. 用户适应互动. , 第 19 卷, no. 3, pp. 243-266, 2009。

X. Xiong, S. Zhao, E. Van Inwegen , 和 J. E. Beck, “用深度知识追踪去更深”, Proc. 第 9 Int.。Conf. Educ. 数据挖掘, 2016, pp. 545-550。

[51] X. Glorot 和 Y. Bengio , “理解训练深度前馈神经网络的困难”, Proc. 13 Int.。Conf. Artif. 情报. 中央集权主义者. , 2010, 第 249-256 页。

董立东, “基于稀疏贝叶斯学习的疫情动态主动监测”, 《中国疾病预防控制中心杂志》第 32 期. 情报. , 2018, pp. 800-807。

A. P. Bradley, “ROC 曲线下面积在机器学习算法评估中的使用”, 《模式识别》. , 第 30 卷, 第 1 期. 7, 第 1145-1159 页, 1997 年。

F.福斯, A.皮罗特, j. m. 和 M. Saerens, “从应用到协作推荐的图节点之间相似度的随机漫步计算”, IEEE 译. 诺尔. 数据工程, 第 19 卷, no. 3, pp 355-369, 2007 年 3 月。



[55] Y. Chen 等人, “跟踪有教育背景的学生的知识熟练程度”, Proc. ACM Conf. Inf. Knowl. 管理。 , 2017, 第 989-998 页。

L. Van der Maaten 和 G. Hinton, “使用 t-SNE 进行数据可视化”, J.马赫。学习。第 9 卷, 第 1 卷。11, 第 2579-2605 页, 2008。

L. M. Rudner, “决策理论适应性测试程序的检验”, 刊于《Proc. Annu》。会见美国。Educ. 辅导班, 2002 年。

[58] Z. Huang 等人, “探索在线教育系统中的多目标运动建议”, 发表于美国计算机学会第 28 届会议。Conf. Inf. Knowl. 管理。 , 2019, 第 1261-1270 页。

[59] H. Zhu 等人, “基于知识地图的多约束学习路径推荐算法”, Knowl. 基于系统。 , vol. 143, pp. 102-114, 2018。

Q. Liu 等人, “开发自适应学习的认知结构”, 发表于第 25 期 ACM SIGKDD Int. Conf. Knowl. 发现。数据挖掘, 2019,pp. 627-635。

[61]张坤, 刘强, 吴瑞华, “基于神经网络的合作学习与比赛结果预测”, 《中国科学(d 科学)》, 2013 年第 1 期。情报。 , 2021, 第 4072-4080 页。

[62]王志强, 王志强, 王志强, 等。“律师执业能力评估网络的构建与应用”, 计算机集成制造系统(cssci), 2003,12(2)。发现。数据挖掘, 2021, 第 5-13 页。



黄振亚, 2014 年毕业于山东大学, 获工学

包括 IEEE 知识与数据工程汇刊、ACM 信息系统汇刊、KDD、AAAI、CIKM 等。

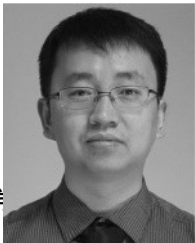


于印, 2017 年毕业于中国科学技术大学计



王飞, 毕业于中国科学技术大学计算机和

, 获工学学士学位。目前在中国科学技术大



王世进, 2003 年毕业于中国科学技术大学



刘琦 (IEEE 成员) , 2013 年毕业于中国科学

博士学位。现任中国科学技术大学计算机科

包括语音处理、自然语言处理和智能教育。带领团队在暴雪挑战赛、CHiME 等国际技术评比中获得十余项冠军。



苏宇, 安徽大学博士。现任合肥师范大

数据和神经计算。他曾获得 KDD'18 最佳学生论文奖和 ICDM'11 最佳研究论文奖。他是阿里巴巴达摩学院青年会员。2019 年获中国杰出青年科学基金。



陈恩宏 (IEEE 高级会员) , 中国科学技术大

科学技术大学计算机科学学院教授、副院长

关于这个或任何其他计算主题的更多信息, 请访问我们的数字图书馆 [www.computer.org/csdl](http://www.computer.org/csdl)。

移动计算、KDD、ICDM、NIPS 和 CIKM 事务处理。他是许多会议的项目委员会成员, 包括 KDD、ICDM、SDM。获 KDD-2008 年度最佳应用论文奖、KDD-2018 年度最佳学生论文奖(研究)、ICDM-2011 年度最佳研究论文奖和 SDM-2015 年度最佳研究论文奖。他的研究受到国家杰出青年科学基金的资助。