

基于 Bloom 认知理论的知识熟练度计算方法^{*}

XXX, XXX, XXX, XXX, XXX

¹(东北大学 计算机科学与工程学院,辽宁 沈阳 110000)

通讯作者: 张天成, E-mail: tczhang@neu.edu.cn

摘 要: 知识熟练度是学生对知识的掌握程度,反映了学生的认知情况。为了准确判断知识熟练度,涌现了多种多样的教育学理论。1956 年提出的 Bloom 认知理论作为经典的理论之一,遵循从底层到高层的认知规律把认知分为多个认知层级,包括“知道”、“领会”“应用”等,构建了层层递进的认知结构。该理论多用于框定教学目标的设计并指引教学活动的发生。本文的研究基于 Bloom 认知理论展开,设计了一种基于 Bloom 认知理论的知识熟练度计算模型 BloomCDM(Bloom Cognitive Diagnosis Model)。该模型基于矩阵分解的思想,具体来说在理论建模中,首先进行层级和层间的理论假设并抽象出数学模型,将题目特征投影到“知道”“领会”“应用”三层认知空间以得到关于目标参数的后验分布。然后,在确定模型的目标函数之后,计算学生和题目特征参数得到学生知识熟练度。在建模过程中,为了抽象出“领会”“应用”的数学表达提出了“知识组”和“高阶知识组”的概念,并设计了高阶知识组结构发现方法。最后,本文的实验证明了该模型可以精准地诊断学生的知识熟练度,证明了在知识熟练度计算中考虑 Bloom 认知层级是科学且有意义的。

关键词: 知识熟练度; 认知诊断; Bloom 认知理论; 矩阵分解

中图法分类号: TP311

中文引用格式: XXXXXXXXXXXXXXX. 基于 Bloom 认知理论的知识熟练度计算方法.软件学报. <http://www.jos.org.cn/1000-9825/0000.htm>

英文引用格式: XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX(in Chinese). <http://www.jos.org.cn/1000-9825/0000.htm>

State-of-the-Art Survey of Static Software Defect Prediction

XXXXXXXXXX

¹(School of Computer Science and Engineering, Northeastern University, Shenyang 110000, China)

Abstract: Knowledge proficiency is the degree to which students master knowledge and reflects the cognitive situation of students. In order to judge knowledge proficiency accurately, a variety of pedagogical theories have emerged. Bloom's cognitive theory proposed in 1956, as one of the classic theories, follows the cognitive law from the bottom to the top and divides cognition into multiple cognitive levels, including "knowing", "understanding" and "application", etc., and constructs a hierarchical cognitive structure. This theory is mostly used to frame the design of teaching objectives and guide the occurrence of teaching activities. The research in this paper is based on Bloom Cognitive theory, and a BloomCDM(Bloom Cognitive Diagnosis Model) for knowledge proficiency calculation is designed based on Bloom Cognitive theory. The model is based on the idea of matrix decomposition. Specifically, in the theoretical modeling, first of all, the hierarchical and inter-hierarchical theoretical assumptions are made and the mathematical model is abstracted. Then, the subject features are projected into the three-layer cognitive space of "knowing", "understanding" and "applying" to obtain the posterior distribution of the target parameters. Then, after determining the objective function of the model, students' and topic characteristic parameters are calculated to obtain students' knowledge proficiency. In the process of modeling, in order to abstract the mathematical expressions of "understanding" and "application", the concepts of "knowledge group" and "higher order knowledge group" are put forward, and the method of finding the structure of higher order knowledge group is designed. Finally, the experiment in this paper proves that the model can accurately diagnose students' knowledge proficiency and that it is scientific and meaningful to consider Bloom's cognitive hierarchy in the calculation of knowledge proficiency.

Key words: Knowledge Proficiency; Cognitive Diagnosis; Bloom's Cognitive Taxonomy; Matrix Factorization

1 引言

随着互联网的快速发展,海量数据的传输和存储得到了广泛的应用。“互联网+教育(Internet + Education)”是近期产生的新型教育模式,在新型教育模式下,“因材施教”依旧是培养高素质高能力人才的重要环节^[1]。由于学生知识熟练度可以反映学

^{*}基金项目: 国家自然科学基金(U1811261)
Foundation item: National Natural Science Foundation of China (00000000, 00000000);
XX
收稿时间: 0000-00-00; 修改时间: 0000-00-00; 采用时间: 0000-00-00; jos 在线出版时间: 0000-00-00
CNKI 在线出版时间: 0000-00-00

生对知识的掌握程度,是学生认知差异的体现。因此精准地计算学生知识熟练度对个性化在线教育尤其重要,可以引导个性化学习空间构建并打造智能化教学环境。

计算知识熟练度一直是教育领域的研究热点^[2],在教育统计领域形成了计算知识熟练度的早期研究体系。早期的认知诊断模型是2009年Torre等人提出了DINA模型^[3]。DINA模型通过离散的二进制向量 $\alpha_i = \{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik}\}$ 来描述学生认知特征,该二进制向量表示学生是否掌握了与题目相关的知识点, $\alpha_{i1} = 1$ 表示学生*i*掌握1知识点,反之亦然。模型设定向量的维度为知识点的数量,并考虑给定的题目—知识点Q矩阵来解释诊断结果。随着数据挖掘方法在知识熟练度计算任务中的运用,提出了一些基于分解模型和深度学习的知识熟练度计算模型,知识熟练度追踪(KPT)^[4]模型改进了传统矩阵分解,在题目特征矩阵V矩阵中添加Q矩阵先验信息,并添加时间窗口进行日志切片来追踪具有解释性的学生知识熟练度U矩阵。基于循环神经网络的深度知识追踪(DKT)^[5]模型采用循环神经网络的隐藏向量表示学生的知识状态,并且据此预测学生的表现。

总结起来,以往的知识熟练度计算多考虑了学生动态的认知变化,添加时间因素,构成三维的追踪模型以构建横向的计算模型。但由于认知的主体是一个复杂体,教育心理学中的Bloom认知理论^[6]已经明确指出认知具备层级特点,真实的认知结构应当是层层递进的,因此在符合教育理论的前提下需构建纵向的知识熟练度计算模型,多角度地量化认知计算。在教育领域,Bloom认知理论主要用于教育目标的设定,体现在题目对不同认知层次的考察。比如:“世界面积最大的大洋是___?”,学生能答出“北冰洋”,可认为该题考察了最底层的“知道”。在某种程度上,题目对更高层次的认知层级有一定的考察,可用来规划试题测试和题目改革。简而言之,当前的研究多在一定程度上构建动态的学习过程追踪模型,但忽略了教育心理学中认知具有层级的特点。文本构建纵向的知识熟练度计算模型,建模在题目对不同认知层次考察下学生知识熟练度的变化。目前,在知识熟练度计算模型中,暂未发现考虑Bloom认知理论来解决知识熟练度计算问题。

为了解决以上问题,本节提出基于Bloom认知理论的知识熟练度计算模型(Bloom Cognitive Diagnosis Model, BloomCDM)。本文的模型从概率矩阵分解中汲取灵感,BloomCDM致力于在题目特征中添加Bloom层级先验,以影响学生认知向量的计算。本文主要创新和贡献有:

- (1) 结合教育心理理论,BloomCDM考虑了“知道”“领会”“应用”三层层级特征和学生知识熟练度特征,BloomCDM根据以上信息建模在题目对不同认知层次考察下学生知识熟练度的变化。BloomCDM将题目特征矩阵投射到每个层次的认知空间并设定为参数 V^R, V^C, V^A ,并根据假设关联学生知识熟练度参数和题目特征参数。
- (2) 为了解决从稀疏数据中获取层级结构信息的问题,基于本文的应用场景,设计了一个高阶知识组发现方法,可以从原始做题矩阵中发现高阶知识组结构并挖掘高阶知识组的结构信息,为基于Bloom认知理论的知识熟练度计算提供了良好的支撑。
- (3) 通过在两个真实在线教育数据集上的实验验证结果表明:BloomCDM可以有效地建模学生的学习行为与遗忘行为,实时追踪学生的知识水平,并且BloomCDM模型的预测性能优于现有模型。

本文第2节介绍知识熟练度计算相关工作。第3节给出相关概念和符号定义,并提出问题。第4节详细介绍BloomCDM模型结构。第5节介绍模型学习和预测的过程。第6节介绍高阶知识组结构发现方法。第7节给出BloomCDM对比实验结果和分析。最后总结全文。

2 相关工作

在本节中,我们将知识熟练度计算研究的相关工作概括为以下三类,分别是认知诊断,矩阵分解和神经网络。

认知诊断作为教育科学中最具挑战性的问题之一,一直受到教育统计领域的广泛关注。认知诊断是被试认知能力高低的判断,在心理测量中占有重要的比重。而认知诊断模型是实现认知诊断测量的重要方法。认知诊断模型主要有连续模型和非连续模型两类。连续模型指的是通过连续变量表示学生潜质,1999年Harvey等人提出的项目反映理论(IRT)^[7]是标准的连续模型,在该模型中,考虑学生一维能力潜变量 θ 和题目特征潜变量来构建学生成绩预测计算模型,并通过Logistic模型预测学生在题目上的表现。2009年Reckase等人提出的多维项目响应理论(MIRT)^[8]是IRT模型的多维版本,考虑了能力 $\theta = (\theta_1, \dots, \theta_m)^T$ 的多维性质,并基于单调性假设和能力独立性假设构建学生成绩预测模型。继经典连续模型之后,在不同的应用场景下多名学者提出了改进的IRT模型,根据人文科学中潜在特质(简称为潜质)具有层次结构的特点,设计了基于线性层级假设计算高低阶潜质变量的HO-CDM^[9]。根据潜质级别的水平结构的特点,将类似潜质水平的被试划分为一个类别以获得被试群体的潜在结构。通常的方法是使用混合Rasch模型(MixIRT)^[10],MixIRT模型得到不易观察到的隐潜质潜在分类,用于评估测量结构的多维性和检测被试作答风格的差异性。非连续模型中最经典的是DINA模型,用多维离散的认知向量建模学生的认知程度,改进的DINA模型有Seq-G-Dina^[11],pG-DINA^[12]等,在题目评分规则是离散多值的情况下,Seq-G-Dina设定受限的掌握水平再结合一般化DINA模型,得到学生离散的认知向量。pG-DINA

像比于G-DINA模型增加了多级知识点的处理，使用多值表示不同程度的水平，分别根据知识点的多级掌握情况计算学生离散认知向量。

矩阵分解是推荐系统中挖掘隐语义关系的一种基础的技术，也是教育数据挖掘常用的算法。主要的应用场景有两类学生表现预测和知识熟练度计算。在学生表现预测中普遍做法是对教育领域中学生成绩数据进行分析，把每个学生映射为学生潜在特征向量。随后进行学生成绩预测，判断学生在题目上的表现。2010年 Thai-Nghe 等人和 Xiong 等人^[13]为了捕捉随时间的推移，学习者知识水平的动态变化，提出考虑时间效应的矩阵分解方法预测学生表现。在学生表现预测任务中，纳入了动态变化的时间维度，从而建模为一个三维（学生、题目、时间）的矩阵分解模型。2016年 P Nedungadi 等人^[14]在预测学生成绩任务中，提出了个性化的加权多关系矩阵分解模型来提高预测精度。模型提出了两种多关系的设计方法，第一种方法考虑全球认知平均水平 μ 和题目偏差项 b 。第二种方法只考虑题目偏差项 b 。综上所述，矩阵分解被证明是非常有效的学生成绩预测模型。该类方法主要通过分解，获得学生的潜在知识向量。但之前工作的一个普遍限制是，这些模型像黑盒一样运行，输出的学生表示很难解释。也就是说，分解得到的隐向量不具备解释性，只能认为是学生的特征向量，若简单的把学生特征向量的维度设为知识点维度就定义为学生的认知向量是比较牵强的。因此，在统计领域，向分解模型中添加可解释性的知识点特征是学生知识熟练度挖掘的新挑战。首次实现基于矩阵分解的知识熟练度计算模型是2017年 Chen 等人^[14]提出 KPT 模型，改进了传统矩阵分解。具体方法是在题目特征矩阵 V 矩阵中添加 Q 矩阵先验信息，其中 Q 矩阵是一个二值的题-知识点矩阵，是认知诊断模型的必要条件。文章改进并利用基于 Q 矩阵的偏序来减少专家标注带来的主观影响。构建偏序集合 D_T 以更准确地捕捉两个知识点 (p, q) 之间的成对关系。实验结果表明 KPT 在多模型多数据的对比中具有出色的效果。随后在 2020 年，Z Huang 等人^[15]基于 KPT 开发了一种改进版本的题目相关的知识熟练度追踪（EKPT）模型来提高预测性能。文章认为 KPT 基于 Q 矩阵的偏序只对每个题目的特征表示单独建模，而忽略了题目之间的关系。因此，在 EKPT 模型中，为对于每个题目 j 定义一个邻居集 N_{V_j} ，邻居集包含与其相似的，具有相同知识点的题目。每个题目 V_j 特征由两项组成，分别是相邻题目的分组特征向量，以及每个题目特征向量，构建了多题目特征影响的动态认知诊断模型。

随着神经网络飞速发展，学生数据和种类趋于完善，基于深度学习的教育数据挖掘成为近期研究的焦点。但由于知识熟练度计

属于可解释性的任务，特征向量需具备认知解释性，和深度学习解释能力弱，可解释力度小的现状背道而驰，因此基于深度学习知识熟练度计算处于萌芽状态，是一个全新的研究领域。近期基于深度学习的学生知识熟练度计算有两个模型：DIRT^[16]和 NeuralCDM^[17]。深度项目反应理论（DIRT）^[16]模型由 Song Cheng 等人于 2019 年提出，该模型认为传统的 IRT 只是利用学生的反应结果进行认知诊断，并没有充分利用题目文本的语义。因此文章提出了一个增强语义的深度学习框架进行认知诊断，通过神经网络模型增强题目文本对题目特征带来的影响，并以 IRT 模型作为损失函数，在大型实际数据集上验证了模型的有效性和解释能力。2020 年，Fei Wang 等人提出神经认知诊断模型（NeuralCDM）^[17]。文章认为现有的方法通常采用 logistic 函数挖掘学生练习过程中的线性交互，这对于捕捉学生与练习之间的复杂关系是不够的。因此文章提出了一个通用的神经认知诊断框架，将学生和题目投射到因子向量上，并利用多神经层来建模，其中应用单调性假设来确保这两个因子的可解释性，综合地考虑了线性和非线性的交互作用。

3 问题定义及预备知识

针对所要进行的具体工作，本节介绍了 BloomCDM 的问题定义以及准备知识。

在教育领域的知识熟练度计算任务中，通常发生在大规模的考试场景中。因此本文的问题定义为：

假设有 N 个学生， M 个题目， K 个知识点，学生集合为 S ，题目集合为 P 。在学生做题日志 R 和教育专家标记的“题目-知识点”矩阵 Q 的条件下，本文的目标是在学生表现预测过程中，建模 Bloom 认知理论下知识熟练度计算模型。可以用公式(1)表示：

$$F(\text{Bloom}, R_i, Q) \rightarrow U_i \quad (1)$$

其中，Bloom 是理论先验， R_i 表示学生 i 的做题记录， Q 表示“题目-知识点”矩阵，是富有教育经验的专家进行标记的。当 $Q_{jk} = 1$ 时，表示题目 j 考察了知识点 k 。当 $Q_{jk} = 0$ 时，表示题目 j 没有考察知识点 k 。 $U_i = \{U_{i1}, U_{i2}, \dots, U_{iK}\}$ 表示学生 i 的知识熟练度向量，向量中的每个值表示学生 i 在知识点 k 上的知识熟练度。从公式(1)可以看出来，研究的重点是设计模型 F ，以原始的评分矩阵和 Q 矩阵的数据作为输入，合理嵌入可解释的 Bloom 先验。不仅在层级建模时，抽象出层级对应的数学模型，还需要纵向地，垂直地构建层间建模，从而得到知识熟练度向量 U_i 。表 1 中列举了问题所需的重要数学符号。

Table 1 Some important notations of BloomCDM

表 1 BloomCDM 主要符号

符号	含义
N	学生总数
M	题目总数
K	知识点总数
H	高阶知识点总数
R_{ij}	学生 i 做题目 j 的结果
U_i	学生 i 的知识熟练度向量
V_j^R	题目 j 的知道向量
V_j^C	题目 j 的领会向量
V_j^A	题目 j 的应用向量
D_T	针对学生的偏序集合
R^C	领会结果矩阵
R^A	应用结果矩阵
CV	子组
PV	父组
P	题目集合

3.1 概率矩阵分解

概率矩阵分解模型 (PMF) [18] 假定真实观测评分 R_{ij} 与预测评分的差值符合均值为 0, 方差为 σ^2 的高斯分布, 即 $P(R_{ij} - U_i^T V_j | 0, \sigma^2)$, 若 $g(x) = 1/(1 + \exp(-x))$ 为 logistic 函数, 假定所有观测评分均相互独立, 可以得到:

$$P(R | U, V, \sigma^2) = \prod_i \prod_j [N(R_{ij} | g(U_i^T V_j), \sigma^2)]^{I_{ij}} \quad (2)$$

假定用户和物品的特征矩阵 U 和 V 满足均值为 0, 方差分别为 σ_u^2 和 σ_v^2 的高斯分布, 同时假定用户特征矩阵 U 中各用户的特征向量满足独立同分布, 物品特征矩阵 V 中各物品的特征向量也满足独立同分布。利用贝叶斯公式, 特征矩阵 U 和 V 的后验分布可通过如下方法计算:

$$\begin{aligned} p(U, V | R, \sigma^2, \sigma_u^2, \sigma_v^2) &\propto p(R_{ij} | U_i^T V_j, \sigma^2) p(U | \sigma_u^2) p(V | \sigma_v^2) \\ &= \prod_{i=1}^N \prod_{j=1}^M N(R_{ij} | U_i^T V_j, \sigma^2)^{I_{ij}} \prod_{j=1}^M N(V_j | 0, \sigma_v^2 I) \prod_{i=1}^N N(U_i | 0, \sigma_u^2 I) \end{aligned} \quad (3)$$

最大化特征矩阵 U 和 V 的后验概率, 等价于最小化上式的负对数。因此, 当超参数 $(\sigma^2, \sigma_u^2, \sigma_v^2)$ 固定时, 可得如下目标函数:

$$J_{PMF} = \sum_{R_{ij} \neq 0} (R_{ij} - g(U_i^T V_j))^2 + \frac{\sigma^2}{\sigma_u^2} \|U\|_F^2 + \frac{\sigma^2}{\sigma_v^2} \|V\|_F^2 + C \quad (4)$$

其中, C 为不依赖于参数的常数。很多成熟的方法都可以用在参数学习中, 比如传统矩阵分解模型适用的梯度类优化方法。同时, 也可依照概率模型求解方法, 即使用最大期望算法 (EM) [19], 马尔可夫链蒙特卡洛算法 (MCMC) [20] 等求解参数 U 和 V 的最大似然估计 (MLE) [21]。

4 基于 Bloom 认知理论的知识熟练度计算模型 (BloomCDM)

针对基于协同过滤的学生个性化试题推荐方法, 本节提出基于 Bloom 认知理论的知识熟练度计算模型 (BloomCDM)。本文的模型从 PMF 中汲取灵感并在题目特征中添加层级先验, 以得到具有良好认知可解释性的知识熟练度向量。

4.1 BloomCDM 整体框架

本节介绍了本文所提出的 BloomCDM 的模型框架, BloomCDM 的整体流程图如图 1 所示。从图中可以看出该推荐方法主要分为三阶段, 分别是数据输入阶段 (Input), 建模阶段 (Modeling) 和输出阶段 (Output)。建模阶段的核心是考虑层级结构 (Hierarchical Structural) 的嵌入包括知道建模, 领会建模和应用建模。

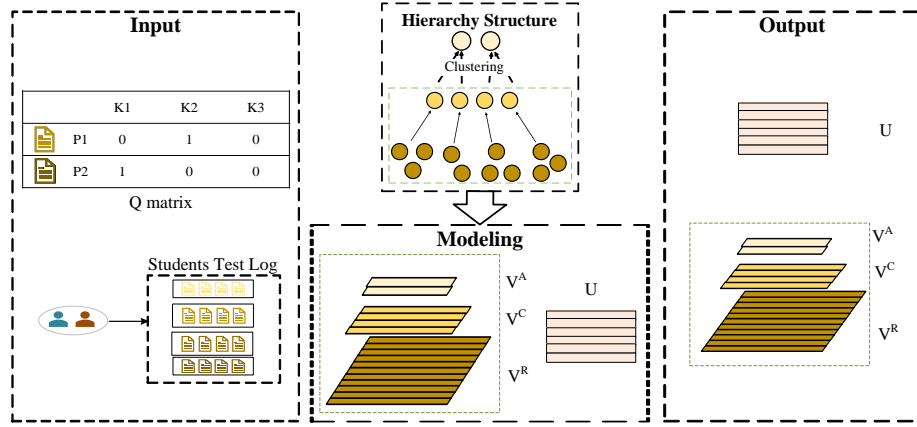


图 1 BloomCDM 结构图

Fig. 1 The Framework of BloomCD model

具体来说, 数据输入阶段提供了学生做题日志和专家标记的题目--知识点矩阵, 分别定义为 R 矩阵和 Q 矩阵。在建模阶段首先进行层级理论假设, 结合 Q 矩阵, 将不同层级的 R 矩阵投影至不同认知空间的认知特征向量。然后进行层间理论假设, 结合模型纵向地, 垂直地, 自下而上地进行层间特征建模。在数据输出阶段, 经过参数估计计算目标参数之后, 输出学生的知识熟练度向量 U 和题目的知道, 领会, 应用特征向量 V^R, V^C, V^A , 实现基于 Bloom 认知理论的知识熟练度计算模型。

4.2 知道建模

知道意味着了解和记忆, 是最低的知识水平。学生不了解知识的含义和内部关系, 而是依靠简单的记忆留下基本印象。

基于知道的特点, 以作对的题目作为研究对象, 假设作对的题目有更大的概率考察知道。也就是说, 既然题目被作对, 相比于没做过的和做错的, 题目一定被测试并考察了基础的知道。那么在知道层级, 作对的题目优先级高于其他题目优先级, 包括做错的和没做过的。

为了符合假设的内容, 本文采用了基于 R 矩阵的偏序^[22], 以突出作对题目的贡献, 强调知道的影响。对于学生 u , 偏序 $>_u^+$ 可定义为公式(5):

$$i >_u^+ j, \text{ 如果 } R_{ui} = 1 \text{ 而且 } R_{uj} = 0 \quad (5)$$

公式(5)表明对于学生 u , 如果一个题目 i 的回答 1, 那么 i 比其他题目更多地考察了知道。在偏序 $>_u^+$ 理论设定的条件下, 可以通过公式(6), 将 R 矩阵变换为一组题目比较性的集合 $D_T \in \mathbb{R}^{N \times M \times M}$:

$$D_T = \{(u, i, j) \mid i >_u^+ j\} \quad (6)$$

因此, D_T 不如 R 矩阵稀疏, 并且可以更精确地捕获基于学生 u 的两个题目 i, j 之间成对关系, 有助于获得对知道的解释。结合偏序先验, 本文设定基于 R 矩阵的知道矩阵 $V^R \in \mathbb{R}^{M \times K}$, 其中 M 表示题目的个数, K 表示知识点的维度。由于偏序基于 R 矩阵, 模型期望得到分解后的学生知识熟练度矩阵 $U \in \mathbb{R}^{N \times K}$ 和题目知道矩阵 V^R , 对所有属于 D_T 的题目对 (i, j) , 通过贝叶斯公式得到后验概率 $p(\theta \mid >_u^+)$ 如公式(7)所示:

$$p(\theta \mid >_u^+) \propto p(>_u^+ \mid \theta) p(\theta) \quad (7)$$

其中, θ 表示模型的参数集合, 也就是 U 和 V^R 参数。那么目标函数分为两个部分。第一部分是模型的似然函数, 第二部分包含 θ 的先验概率。对于第一部分 $p(>_u^+ \mid \theta)$ 而言, 在知道层级, 同时假设题目对 (i, j) 独立于其他题目对。因此, 似然函数 $p(>_u^+ \mid \theta)$ 可得公式(8):

$$p(>_u^+ \mid \theta) = \prod_{(u, i, j) \in D_T} p(i >_u^+ j \mid \theta) \quad (8)$$

为了得到在 θ 条件下正确的偏序关系, 定义学生 u 比题目 i 与题目 j 更相关的概率为公式(9):

$$p(i >_u^+ j \mid \theta) = \sigma(\bar{x}_{ui}(\theta) - \bar{x}_{uj}(\theta)) \quad (9)$$

其中 σ 是 logistic 函数, 为了同时满足偏序 $i >_u^+ j$ 的完整性, 反对称性, 传递性和优化计算而设计的。因为 logistic 函数本身是连续且中心对称函数, 满足偏序的设定。此外, 对于第二部分 $p(\theta)$ 而言, 依据传统的处理, 假设 U 和 V^R 均满足高斯分布, 因此 θ 遵循一个零均值的高斯先验。结合目标函数的第一部分公式(8), (9)和第二部分的高斯假设, 可以得到 D_T 条件下 θ 的后验分布为公式(10):

$$p(\theta | >_u^+) \propto p(>_u^+ | \theta) p(\theta) = \prod_{(u,l,j) \in D_T} p(>_u^+ | V^P) p(V^P) \quad (10)$$

4.3 领会建模

领会意味着理解事物，但不需要深入理解，相比于知道，领会不仅仅是简单地考察记忆知识点，进一步考察对浅层概念和意义上的把握。

基于领会的特点，以考察同一知识点的题目作为研究对象。假设相同知识点都被考察时，能更进一步测量该知识点的理解度。也就是说，当仅考察该知识点的题目时，对于该知识点的内在逻辑和形式都会有一定程度的印象，加深对该知识点的理解程度。

因此我们提出**知识组**的概念：形成**知识组**的组织结构。依据 Q 矩阵所提供的题目-知识点类别标签，将考察同一知识点的题目划分为不重叠，不交叉，不相交的组，从下至上构建**知识组**结构的理解模型。和 Q 矩阵并不相同，知识组虽源于 Q 矩阵，但规定了基础的认知计算单元，将扁平稀疏的 Q 矩阵层级化，方便科学计算。知识组的提出，不仅是为了捕获题目之间相同知识相关的特征，同时为了更好的搭建层级结构。接下来详细介绍一些相关概念：知识组，子组 CV_l^C ，父组 PV_l^R ，领会矩阵 V^C ，评分矩阵 R^C 。

知识组：定义属于相同知识点的题目构成了一个该题目的知识组。例如，在图 1 中，有 7 个学生，11 个题目，4 个知识组 $\{1^{(V^C)}, 2^{(V^C)}, 3^{(V^C)}, 4^{(V^C)}\}$ 。层级结构模块中第一行实心圆圈表示题目，第二行实心圆圈表示知识组。题目 1、2、3 属于 $1^{(V^C)}$ 知识组，题目 4、5 属于 $2^{(V^C)}$ 知识组，题目 6、7、8 属于 $3^{(V^C)}$ 知识组，题目 9、10、11 属于 $4^{(V^C)}$ 知识组。根据本文的任务，拟定知识组具备以下特点：

- 每一个知识组内元素有且仅归属一个知识点。
- 组内至少一个元素。
- 组间无重叠。

子组 CV_l^C ：定义 CV_l^C 为知识组 l 的题目集合，例如，题目 1、2、3 属于 $1^{(V^C)}$ 且是知识点 1 的子组， $CV_1^C = \{1, 2, 3\}$ ，其余子组依次类推。

父组 PV_l^R ：题目 l 可以通过 PV_l^R 映射到对应的知识组。例如，题目 1、2、3 属于 $1^{(V^C)}$ 因此父组 $PV_1^R = 1$ 。其余父组依次类推。

此处，领会矩阵 $V^C \in \mathbb{R}^{K \times K}$ 表示领会矩阵，表示题目对于领会的考察程度， K 表示知识组的数量，同时 K 也是特征向量的维度。因为 R 矩阵没有知识组特征，参数将很难学习到领会方面的内容。因此本文引入领会层级结果矩阵 R^C ，有原则的关联知识组和结果矩阵 R ，以进一步监督 V^C 的训练。

本质上来讲， R^C 是一个学生对知识组的评分矩阵，维度是 $R^C \in \mathbb{R}^{N \times K}$ ，其中 N 表示学生的数量， K 表示知识组的数量， R_{uj}^C 表示学生 u 在知识组 j 的做题结果。 R^C 可以用多种方式构造，比如当各组数据在知识组内呈均匀分布时，使用组中值或选取最大，最小值作为知识组的代表值。由于知识组内做题数据通常是二值的，也存在一定比例的缺值，因此模型采用组内平均作为知识组的代表值。知识组结果矩阵的计算公式(11)：

$$R_{uj}^C = \frac{1}{|CV_j^C|} \sum_{l \in CV_j^C} R_{ul} = \bar{R}_{u, CV_j^C} \quad (11)$$

在建立学生，题目和知识组之间的联系时，由于题目特征矩阵具有层级特点，这几个层构成了累积的层次框架，达到下一个层次需在前一层次的基础上。因此假设每一层题目特征矩阵从它们的子特征矩阵中采样，针对知道和领会， $V^C \sim \mathcal{N}(V^R, \sigma_V^2 I)$ 。那么生成过程如下：

- (1) 针对知道和领会的层级特点，有： $V^C \sim \mathcal{N}(V^R, \sigma_V^2 I)$
- (2) 有 $U \sim \mathcal{N}(0, \sigma_U^2 I)$
- (3) 对于领会层级每个未丢失的 (u, j) ，有 $R_{uj}^C \sim \mathcal{N}(\langle U_u, V_j^C \rangle, \sigma_R^2 I)$ 。
- (4) δ 为指示函数，如果 $R_{uj}^C \neq null$ 时， $\delta_{uj}^R = 1$ ，否则 $\delta_{uj}^R = 0$

其中 σ 为标准差，则对参数 V^R, V^C 和 U 的后验分布为公式(12)：

$$\begin{aligned} & P(V^R, V^C, U | >_u^+, R^C, \sigma_R^2, \sigma_U^2, \sigma_V^2) \\ & \propto P(>_u^+ | U, V^R) P(R^C | U, V^C, \sigma_R^2) P(V^C | \sigma_V^2) P(U | \sigma_U^2) P(V^R | \sigma_V^2) \\ & = \prod_{(u,l,j) \in D_T} \sigma(\bar{x}_{ui} - \bar{x}_{uj}) \\ & \quad \prod_{u,j} \mathcal{N}(R_{uj}^C | \langle U_u, V_j^C \rangle, \sigma_R^2) \delta_{uj}^{R^C} \end{aligned}$$

$$\mathcal{N}(V^C | V^R, \sigma_V^2 I) \mathcal{N}(V^R | 0, \sigma_V^2 I) \mathcal{N}(U | 0, \sigma_U^2 I) \quad (12)$$

4.4 应用建模

应用是指对所学习的概念、法则、原理的运用。相比于领会，应用不仅仅是考察对该知识点的了解和把握程度，更重要的是通过该知识点的了解，获得一种举一反三的能力，去学习其他的知识点，是一种跨知识点的能力。

基于应用的特点，以考察相似知识点的题目作为研究对象。假设相似知识点都被考察时，更容易发现知识之间的内在联系。也就是说，当题目考察相似知识点时，更容易测量概念或知识点领会的迁移水平，能进一步加深对知识概念的理解程度。

因此本文创新性地提出了**高阶知识组**的概念，形成高阶知识组的组织结构，构建跨知识点的计算模型。和领会类似，将考察相似知识点的知识组划分为不重叠，不交叉，不相交的组（簇），其中，由相似知识点抽象出高阶知识组，不仅简洁地描述了知识组的聚合程度，更能捕获相似知识组相关的特征，更好的搭建层级结构。高阶知识组基于知识组，是知识组更高抽象的表示。和知识组类似，存在一些相关概念：高阶知识组，子组 CV_l^A ，父组 PV_l^C ，应用矩阵 V^A ，应用评分矩阵 R^A 。

高阶知识组：相似知识组构成了高阶知识组。例如，在图 1 中，存在 11 个题目，4 个知识组 $\{1^{(V^C)}, 2^{(V^C)}, 3^{(V^C)}, 4^{(V^C)}\}$ 和 2 个高阶知识组 $\{1^{(V^A)}, 2^{(V^A)}\}$ 。设定，题目 $\{1, 2, 3\}$ 属于 $1^{(V^C)}$ 知识组，题目 $\{4, 5\}$ 属于 $2^{(V^C)}$ 知识组，题目 $\{6, 7, 8\}$ 属于 $3^{(V^C)}$ 知识组，题目 $\{9, 10, 11\}$ 属于 $4^{(V^C)}$ 知识组。由于高阶知识组的提出，因此产生了知识组到高阶知识组的对应关系。如图，知识组 $\{1^{(V^C)}, 2^{(V^C)}\}$ 属于高阶知识组 $1^{(V^A)}$ ，知识组 $\{3^{(V^C)}, 4^{(V^C)}\}$ 属于高阶知识组 $2^{(V^A)}$ 。高阶知识组的特点和知识组的特点类似，此处不再赘述。

子组 CV_l^A ：定义 CV_l^A 为高阶知识组 l 的知识组集合，例如，知识组 1、2 属于 $1^{(V^A)}$ ， $CV_1^A = \{1, 2\}$ ，其余子组依次类推。

父组 PV_l^C ：知识组 l 可以通过 PV_l^C 映射到对应的高阶知识组。例如，知识组 1、2 属于 $1^{(V^A)}$ ，因此父组 $PV_1^C = 1$ 。其余父组依次类推。

此处， $V^A \in \mathbb{R}^{H \times K}$ 表示应用矩阵， H 表示高阶知识组的个数。和领会类似， V^A 依旧是通过带监督的训练得到，因此本文还需引入应用层级结果矩阵 R^A ，以进一步监督 V^A 的训练。 R^A 大小是 $\mathbb{R}^{N \times H}$ ， R_{uj}^A 表示学生 u 在高阶知识组 j 的做题结果。 R^A 可以用多种方式构造，模型依旧采用组内平均作为组的代表值。高阶知识组结果矩阵的计算公式(13)：

$$R_{uj}^A = \frac{1}{|CV_j^A|} \sum_{l \in CV_j^A} R_{ul}^C = \bar{R}_{u, CV_j^A} \quad (13)$$

需要注意的一点是，在大多数应用场景中，**高阶知识组**结构信息是隐式的，很难直观地获得。结构的发现方法本文将在第四章进行论述，此处可认为通过下一节的高阶知识组结构发现算法获得高阶知识组结构信息，并不是在模型训练的过程中获得的。

在建立学生、题目和知识组之间的联系时，由于题目特征矩阵具有层级特点，这几个层构成了累积的层次框架，在前一层次的基础上达到下一个的层次。因此假设每一层的题目特征矩阵从它们的子特征矩阵中采样，针对领会和应用来讲， $V^A \sim \mathcal{N}(V^C, \sigma_V^2 I)$ 。那么生成过程如下：

- (1) 针对领会和应用的层级特点，有： $V^A \sim \mathcal{N}(V^C, \sigma_V^2 I)$
- (2) 有 $U \sim \mathcal{N}(0, \sigma_U^2 I)$
- (3) 对于应用层级每个未丢失的 (u, j) ，有 $R_{uj}^A \sim \mathcal{N}(\langle U_u, V_j^A \rangle, \sigma_R^2 I)$ 。
- (4) δ 为指示函数，如果 $R_{uj}^A \neq null$ 时， $\delta_{uj}^A = 1$ ，否则 $\delta_{uj}^A = 0$

其中 σ 为标准差。则参数 V^R, V^C, V^A 和 U 在偏序和每层的结果矩阵的条件下的后验分布为公式(14)：

$$\begin{aligned} & P(V^R, V^C, V^A, U | >_u^+, R^C, R^A, \sigma_R^2, \sigma_R^2, \sigma_U^2, \sigma_V^2) \\ & \propto P(U, V^R | >_u^+) P(R^C | U, V^C, \sigma_R^2) P(R^A | U, V^A, \sigma_R^2) P(U | \sigma_U^2) P(V^R | \sigma_V^2) P(V^C | \sigma_V^2) P(V^A | \sigma_V^2) \\ & = \prod_{(u, i, j) \in D} \sigma(\bar{x}_{ui} - \bar{x}_{uj}) \\ & \prod_{u, j} \mathcal{N}(R_{uj}^C | \langle U_u, V_j^C \rangle, \sigma_R^2) \delta_{uj}^{R^C} \\ & \prod_{u, j} \mathcal{N}(R_{uj}^A | \langle U_u, V_j^A \rangle, \sigma_R^2) \delta_{uj}^{R^A} \\ & \mathcal{N}(V^C | V^R, \sigma_V^2 I) \mathcal{N}(V^A | V^C, \sigma_V^2 I) \mathcal{N}(V^R | 0, \sigma_V^2 I) \mathcal{N}(U | 0, \sigma_U^2 I) \end{aligned} \quad (14)$$

5 模型学习和预测

为了更好地理解提出的 BloomCDM，本节在图 2 中总结了模型的概率图表示，其中阴影和无阴影变量表示观察到的和潜在的变量。

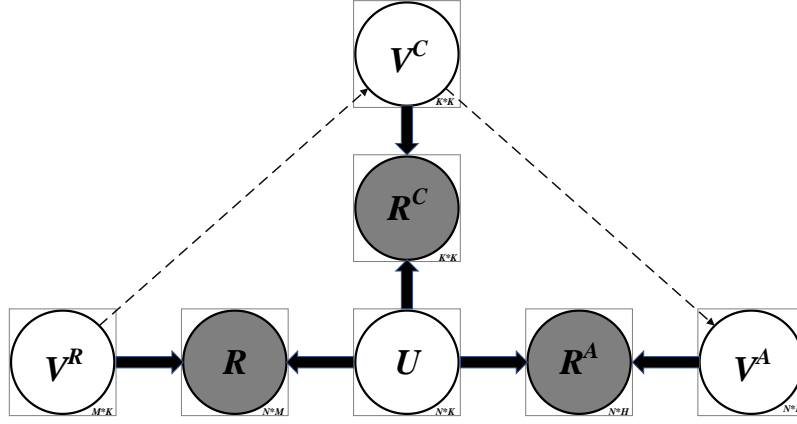


图 2 BloomCDM 概率图

Fig. 2 Graphical model of BloomCDM

通过上节的详细描述，已知学生做题结果 R 矩阵，领会层级知识组 R^C 矩阵，应用层级高阶知识组 R^A 矩阵和偏序 $>_u^+$ ，模型的目标是学习参数 $\phi = [U, V^R, V^C, V^A]$ 。通过 BloomCDM 模型，得到了在偏序和每层的结果矩阵的条件下目标参数 ϕ 的后验分布。

因此，公式(14)的对数后验最大化等价于使公式(15)目标函数最小化：

$$\begin{aligned}
 \mathcal{L} &= \ln p(V^R, V^C, V^A, U | >_u^+, R^C, R^A, \sigma_R^2, \sigma_U^2, \sigma_V^2) \\
 &\propto - \sum_{(u,l,j) \in D_T} (\ln \sigma(U_u \cdot V_l^R - U_u \cdot V_j^R)) \\
 &\quad - \frac{1}{2\sigma_R^2} \sum_{u=1}^N \sum_{j=1}^K \delta_{uj}^{R^C} (R_{uj}^C - \langle U_u, V_j^C \rangle)^2 \\
 &\quad - \frac{1}{2\sigma_R^2} \sum_{u=1}^N \sum_{j=1}^H \delta_{uj}^{R^A} (R_{uj}^A - \langle U_u, V_j^A \rangle)^2 \\
 &\quad - \frac{1}{2\sigma_V^2} \sum_{j=1}^M (V_j^R - V_{PV_j^R}^C)^2 \\
 &\quad - \frac{1}{2\sigma_V^2} \sum_{j=1}^K (V_j^C - V_{PV_j^C}^A)^2 \\
 &\quad - \frac{1}{2\sigma_V^2} \sum_{j=1}^M (V_j^R)^2 - \frac{1}{2\sigma_U^2} \sum_{u=1}^N U_u^2
 \end{aligned} \tag{15}$$

机器学习参数求解最常用的三种方法：最小二乘法，梯度下降法，最速牛顿法。由于目标函数是凸函数，模型选用随机梯度下降法(SGD)^[23]。为了更准确地收敛，模型并不是直接应用 SGD，而是遵循随机块坐标下降在每个级别上依次更新 U 和所有的 V 。

$$\frac{\partial \mathcal{L}_R}{\partial U_u} = \frac{1}{1 + e^{U_u \cdot V_l^R - U_u \cdot V_j^R}} (V_l^R - V_j^R) - \frac{1}{\sigma_U^2} U_u \tag{16}$$

$$\frac{\partial \mathcal{L}_{R^C}}{\partial U_u} = \frac{1}{\sigma_R^2} \sum_{j=1}^K \delta_{uj}^{R^C} (R_{uj}^C - \langle U_u, V_j^C \rangle) V_j^C - \frac{1}{\sigma_U^2} U_u \tag{17}$$

$$\frac{\partial \mathcal{L}_{R^A}}{U_u} = \frac{1}{\sigma_R^2} \sum_{j=1}^H \delta_{uj}^{R^A} (R_{uj}^A - \langle U_u, V_j^A \rangle) V_j^A - \frac{1}{\sigma_U^2} U_u \quad (18)$$

$$\frac{\partial \mathcal{L}_R}{V_j^R} = \frac{U_u}{1 + e^{U_u \cdot V_j^R - U_u \cdot V_j^R}} - \frac{1}{\sigma_V^2} V_j \quad (19)$$

$$\frac{\partial \mathcal{L}_R}{V_j^C} = \frac{1}{\sigma_R^2} \sum_{i=1}^K \delta_{ij}^{R^C} (R_{ij}^C - \langle U_i, V_j^C \rangle) U_i - \frac{1}{\sigma_V^2} \sum_{i=1}^K [|CV_j^C| (V_j^C - \bar{V}_{CV_j^C}^R) + (V_j^C - V_{PV_j^C}^A)] \quad (20)$$

$$\frac{\partial \mathcal{L}_R}{V_j^A} = \frac{1}{\sigma_R^2} \sum_{i=1}^H \delta_{ij}^{R^A} (R_{ij}^A - \langle U_i, V_j^A \rangle) U_i - \frac{1}{\sigma_V^2} \sum_{i=1}^H [|CV_j^A| (V_j^A - \bar{V}_{CV_j^A}^C)] \quad (21)$$

因此, 算法 1 给出了 BloomCDM 评分预测的学习过程。可以看到时间复杂度和阶数线性相关, 本文设定阶级为 3, 假设在所有结果矩阵中有非空项, 每个学生的平均时间为 $tr = r/N$, 在每次迭代中, U 的时间复杂度为 $O(N \times tr) = O(K \times r)$, V 的时间复杂度为 $O(3 \times M \times tr) = O(K \times r)$ 。因此, 每次迭代中参数学习的总复杂度为 $O(K \times r)$, 与记录和呈线性关系。

算法 1: BloomCDM 的参数学习算法

Algorithm 1: The learning algorithm of BloomCDM

输入: 矩阵 R , 子组 CV^C , CV^A , 标准差 σ_U^2 , σ_V^2 , σ_R^2 , 学习率 η , 迭代次数 T , 层级层数 $L=3$ 。

输出: 学生特征矩阵 U , 层级特征矩阵 V^P, V^C, V^A , 预测的 \tilde{R} 矩阵。

```

1. for t = 1, 2, ..., T do
2.   for l = 1, 2, ..., L do
3.     if l = 1 do
4.        $U \leftarrow U + \eta [\partial \mathcal{L}_R / U_u]_{u=1}^N$  via Eq. (16)
5.        $V^R \leftarrow V^R + \eta [\partial \mathcal{L}_R / V_j^R]_{j=1}^M$  via Eq. (19)
6.     end if
7.     if l = 2 do
8.        $U \leftarrow U + \eta [\partial \mathcal{L}_{R^C} / U_u]_{u=1}^N$  via Eq. (17)
9.        $V^C \leftarrow V^C + \eta [\partial \mathcal{L}_{R^C} / V_j^C]_{j=1}^K$  via Eq. (20)
10.    end if
11.    if l = 3 do
12.       $U \leftarrow U + \eta [\partial \mathcal{L}_{R^A} / U_u]_{u=1}^N$  via Eq. (18)
13.       $V^A \leftarrow V^A + \eta [\partial \mathcal{L}_{R^A} / V_j^A]_{j=1}^H$  via Eq. (21)
14.    end if
15.  end for
16. end for
17. return  $\tilde{R} = \langle U_u, V_j^R \rangle$      $\tilde{R}^C = \langle U_u, V_j^C \rangle$      $\tilde{R}^A = \langle U_u, V_j^A \rangle$  //输出特征矩阵和预测矩阵

```

6 高阶知识组结构发现

在之前提到, 应用层级是通过对单一知识点的领会, 获得一种举一反三的能力, 去学习其他的知识点, 是一种跨知识点的能力。因此提出“高阶知识组”的概念, 构建跨知识点的计算模型。但在某些应用场景中, 这种层级结构信息是隐式的, 可能很难直观的获得结构信息。我们提出了一种高阶知识组结构发现方法, 从知识组中抽象出高阶知识点, 进而完善高阶知识组结构的构建, 为基于 Bloom 认知理论的知识熟练度计算提供了良好的支撑。

由于学生反馈的稀疏性, 通常很难设计出一种合适的聚类算法有效地将相关知识组归为同一组。根据两种直观的假设, 第一个是直接应用聚类算法在知识组中, 另一种直觉根据高维且稀疏的数据特点, 提出了先知识组特征降维, 后聚类算法两阶段的研究路线。

根据直觉一, 我们首先在聚类场景下选用 DBSCAN 算法发现高阶知识组结构。DBSCAN 错误!未找到引用源。 是一种聚类算法, 其不要求指定簇的数量, 特别适合高阶知识组的发现。DBSCAN 算法的思想是在每个数据点周围创建一个 ϵ 半径的圆, 将所有数据点分类为核心点, 边界点和噪音。在这个半径内拥有超过 minPts 邻居的数据点被认为是一个核心点。在核心点的附近 minPts 的数据点都被称为直接密度可达点, 直接密度可达点若是被认定为核心点, 那么其邻居点被定义为密度可达, 同一个簇中的所有点都是密度

连接的。如果点数小于 minPts 被认为是一个边界点，周围没有其他数据点，则将其视为噪音。直接应用 DBSCAN 的聚类效果如图 3(左)所示。可以从图中观察到直接应用的情况下，DBSCAN 算法无法获得聚类个数且没有挖掘知识组的相似特性。

按照直觉二，先对知识组特征进行降维，随后 DBSCAN 进行聚类。在降维阶段，t-分布随机邻居嵌入算法(t-SNE)是 2008 年提出的重要降维方法，模型的初衷是可视化高维数据，在很多领域都有广泛的应用。t-SNE 算法第一步计算高维空间数据点之间的互为邻域的成对距离的概率分布，第二步计算低维空间数据点之间的互为邻域的成对距离的概率分布。第三步时，算法试图最小化高维和低维空间中这些条件概率（或相似性）之间的差异，以实现低维空间中数据点的完美表示。KL 散度度量高维和低维数据的分布之差，为了使得 KL 散度最小化，由于凸函数的性质，t-SNE 使用梯度下降法最小化两个分布的距离。总而言之在知识组特征降维应用场景下，t-SNE 最小化两个分布的差异，分别是知识组的高维特征分布在低维空间中的分布。在聚类阶段，依旧通过 DBSCAN 对低维数据点进行聚类发现。直觉二的聚类效果如图 3(右)所示，可以清晰地分辨不同的簇，获得簇的个数为 3，聚类模块度更高较为精准地捕捉知识组之间的关系。

算法 2：高阶知识组结构发现

Algorithm 2: Higher-order Knowledge Group Structure Detection

输入： 领会结果矩阵 $R^C = \{R_1^C, R_2^C, \dots, R_n^C\}$ ，困惑度 $\text{Perp}(i)$ ，迭代次数 T ，学习速率 η ， ϵ 半径， minPts

初始化 label 集为 undefined

输出： 高阶知识组的个数，label 集

```

1.  $DB = \text{t\_SNE}(R^C, \text{Perp}, T, \eta)$            // 通过t_SNE降维，得到降维后的低维数据点
2. for 每一个数据点  $p$  in  $DB$  do
3.   if  $\text{label}(p) \neq \text{undefined}$  then continue   // 选取一个未处理的点
4.   邻居  $N \leftarrow \text{RangeQuery}(T, p, \epsilon)$        // 寻找 $p$ 点密度可达的点，并添加到邻居 $N$ 中
5.   if  $|N| < \text{minPts}$  then                       // 如果邻居点的数量小于 minPts 时， $p$ 点暂被标记为噪音
6.      $\text{label}(p) \leftarrow \text{噪音}$ 
7.     继续
8.    $c \leftarrow$  下一个聚类label
9.    $\text{label}(p) \leftarrow c$ 
10.  邻居集  $S \leftarrow N\{p\}$ 
11.  for 每一个数据点  $q$  in  $S$  do
12.    if  $\text{label}(q) = \text{噪音}$  then  $\text{label}(q) \leftarrow c$ 
13.    if  $\text{label}(q) \neq \text{undefined}$  then continue
14.    邻居  $N \leftarrow \text{RangeQuery}(T, q, \epsilon)$ 
15.     $\text{label}(q) \leftarrow c$ 
16.    if  $|N| < \text{minPts}$  then continue
17.     $S \leftarrow S \cup N$ 
18.  end for
19. end for
20. return label, unique(label)

```

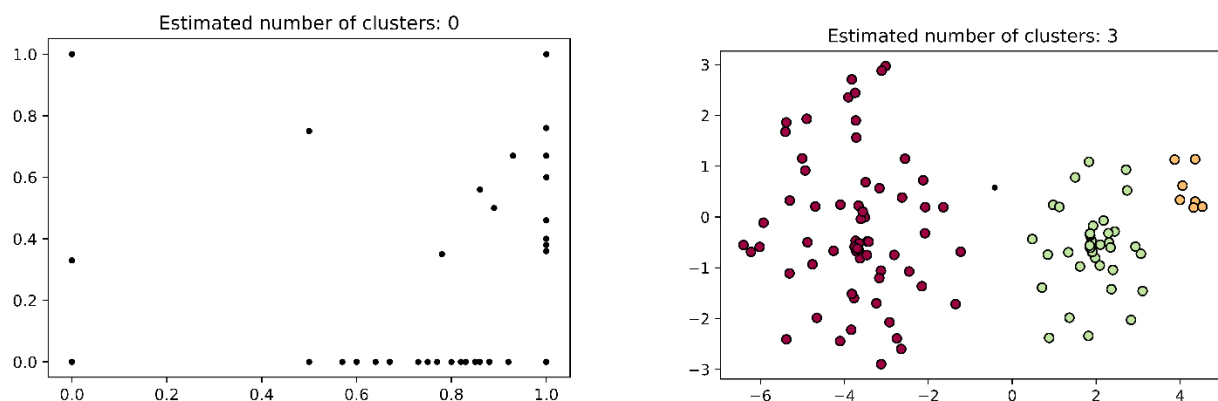


图3 聚类方法对比
Fig. 3 The comparison of clustering method

7 实验

在本节中, 我们首先介绍我们的实验数据集和设置。然后, 我们从以下三个方面报告实验结果: (1) 学生表现预测任务; (2) 知识熟练度诊断任务; (3) 可视化知识熟练度任务。

7.1 数据集

本实验是研究在做题场景下学生的认知情况, 因此需要学生在题目上的表现数据, 最重要的是获得学生的做题得分数据。目前在认知诊断领域的很多研究, 都是依托教育机构和组织所提供的私有数据集, 彻底公开的数据集比较稀少。因此在本文的实验中, 使用了真实世界中的两个数据集, 分别是: ASSIST 数据集和杭州电子科技大学 OJ 平台数据集 (HDU)。

ASSIST 数据集源于公开数据集: Assistments 2009-2010 年 “skill builder”¹。公开数据集是由 Heffernan 教授和他的团队在国外的在线教育平台 Assistments 上整理的数据集。在公开数据集中, 详细记录了 2009-2010 年间学生的做题日志, 包括学生序号, 题目序号, 做题结果, 教育模式, 学生的尝试数等, 并标注了题目对应的技能项, 比如 Percent Of (百分比), Absolute Value (绝对值) 等。HDU 数据集源于杭州电子科技大学 OJ 平台。OJ 平台是在线判题系统, 通过发布题目, 平台收集用户的做题数据, 进行自主编译和运行, 验证学生做题的正确性。OJ 平台的特点是对所有用户开放, 只要登录系统就可以提交代码进行测评。国内比较出色的 OJ 平台有 HDU (杭州电子科技大学), PKU (北京大学) 等。在 HDU 的 OJ 平台的 [Realtime Judge Status] 模块中, 有大量用户做题历史日志。本实验爬取数据的目标字段有: “Author”, “ProID”, “Judge Status”, “Submit Time”, 爬取范围是 2020-2021 年学生做题日志。题目所对应知识点类别主要有: 图论, 树论, 深度优先搜索, 排序等, 知识点的确定和标注均由实验组的同学共 15 人协作完成。为了解决本文提出的任务, 本实验实施几点预处理操作:

(1) 去重操作。由于两个数据集均具有时序关系, 即在同一个题目上, 一个学生会有多次的回答记录, 因此做题日志是序列数据。本实验只保留第一次的回答记录, 认为学生第一次回答是学生真实认知的反馈以保证学生做题结果的唯一性。

(2) 解决长尾分布问题。由于两个数据集存在长尾分布, 其中做题日志过少, 做题不活跃的学生易影响诊断结果。本实验过滤了做题日志小于 15 的学生和学生日志小于 15 的题目, 以确保每个学生和题目都有足够的数据诊断学生的认知情况。经过过滤和整理的数据结构如表 3.2 所示。

(3) 统一做题状态。由于 HDU 数据集的做题状态有 4 种, “编译错误, 超时, 错误答案, 成功”, 本实验剔除了 “编译错误, 超时” 两种状态, 仅考虑 “错误答案, 成功” 这两种状态。

7.2 实验设置

BloomCDM 设置: 为了评估 BloomCDM 模型的性能, 本实验将数据集按 80% 的比例划分为训练数据集和测试数据集, 对不同数据稀疏度下的性能预测任务进行了研究。针对 BloomCDM 模型中的超参数, 默认的参数设置如下:

- 学习率 η : 0.001, 标准差 σ_U^2 , σ_V^2 , σ_R^2 : 0.001

针对高阶知识组发现方法中的超参数设置如下:

¹ <https://sites.google.com/site/assistmentsdata/home/assistent-2009-2010-data>

● Perp(i): 30, 学习率 η : 50, ε 半径: 2, minPt: 21, Dist:

基准方法: 为了比较 BloomCDM 模型的性能, 验证模型的健壮性, 本文选用了 6 种基准方法进行了对比试验。

(1) IRT

项目响应理论 IRT 是经典的教育统计学中的认知诊断模型。如公式(22)所示。考虑学生一维能力潜变量 θ 和题目特征潜变量 d 来构建学生反应概率计算模型。

$$p_{ij} = \sigma(\theta_i - d_j)$$
 (22)

(2) MIRT

多维项目响应理论 MIRT 是 IRT 模型的多维版本, 如公式(23)所示。考虑能力 $\theta = (\theta_1, ..., \theta_m)^T$ 的多维性质, 并基于单调性假设和能力独立性假设构建学生反应概率计算模型。

$$p_{ij} = \sigma(\langle \theta_i, d_j \rangle + \delta_j)$$
 (23)

(3) PMF

概率矩阵分解 (PMF) 是一个应用非常广泛的推荐系统算法, 采用因式分解方法, 将做题日志分解为学生 U 和题目 V 的隐特征矩阵。

(4) QPMF

QPMF^[4]是 PMF 的变体, 引入 Q 矩阵以提高 PMF 的解释力。嵌入方法是采用基于 V 矩阵的偏序, 以突出题目考察知识点的贡献。

(5) BPR

贝叶斯个性化推荐 (Bayesian Personalized Ranking, BPR) 是推荐系统的经典算法, BPR 结合了基于偏序关系构造的似然函数 $P(i > uj | \Theta)$ 和先验概率 $P(\Theta)$, 对学生做题日志进行贝叶斯分析。

(6) BloomCDM-RC

BloomCDM-RC 是 BloomCDM 的简化版本。仅仅考虑知道 (Remember) 和领会 (Comprehension), 没有考虑应用层级。具体来说, 所选择的基线均广泛应用于传统认知诊断(IRT、MIRT), 推荐系统模型(PMF、BPR、QPMF)和自身的模型的一个变体(BloomCDM-RC)。在接下来的实验中, BloomCDM 和基线都是由 Python 实现的, 所有的实验都运行在 Linux 服务器上, 有 4 个 2.0GHz Intel Xeon E5-2620 cpu 和 100G 内存。为了公平起见, 这些基线中的所有参数都被调整为具有最佳性能。

7.3 实验结果

学生表现预测结果分析

认知诊断模型的性能很难评价, 因为我们不能获得学生的真实知识水平。在大多数研究中, 诊断结果通常是通过预测学生的表现来获得的, 这些预测任务的表现可以从一个方面间接评价模型。考虑到我们在数据中使用的所有题目都是客观题, 采取之前设定的超参数我们从分类方面和回归方面使用评价指标, 因此采用了回归模型最常见的两个指标: 均方根误差 (RMSE), 平均绝对误差 (MAE) 和分类模型最常见的两个指标: 准确率 (ACC), ROC 曲线下面积 (AUC)。表 2 为各模型对学生成绩预测任务的实验结果。

Table. 2 Comparative experimental results of all models in student performance prediction tasks
表 2 所有模型在学生表现预测任务的对比实验结果

模型	ASSIST				HDU			
	RMSE	MAE	ACC	AUC	RMSE	MAE	ACC	AUC
IRT	0.463	0.398	0.648	0.648	0.533	0.398	0.676	0.625
MIRT	0.450	0.386	0.750	0.678	0.471	0.386	0.736	0.750
PMF	0.460	0.394	0.657	0.657	0.479	0.394	0.724	0.657
QPMF	0.451	0.388	0.674	0.683	0.460	0.397	0.744	0.687
BPR	0.449	0.386	0.678	0.750	0.449	0.366	0.722	0.678
BloomCDM-RC	0.422	0.370	0.785	0.785	0.412	0.370	0.754	0.785
BloomCDM	0.421	0.364	0.836	0.886	0.407	0.364	0.766	0.836

综合所有数据集的分类和回归实验结果, 可以得到: (a) 无论是分类还是回归, BloomCDM 效果优于传统认知诊断模型 IRT, MIRT 和推荐模型 BPR, PMF。说明本文的设计是科学有效的, 可以提升模型学生表现预测效果。(b) BloomCDM 效果优于 BloomCDM_RC, 特别是 ACC 指标中, BloomCDM 相比于 BloomCDM_RC 提升 0.5 个百分点左右, 表明应用层级的建模是不可或缺的, 需要捕捉相似知识点的特征提升预测能力。(c) 传统认知诊断模型 MIRT 在 AUC 指标上效果较好, 甚至比推荐模型 BPR, PMF 效

果要好, 这是个很有趣的现象。本文认为随后的模型的研究点可以是在 MIRT 模型上添加知识点, 来计算更为精确的知识熟练度。

(d) 在 AUC, ACC 指标中, ASSIST 数据集的效果比 HDU 数据集的效果好, 可能是 OJ 平台的模式是自由练习, 可能存在更多的隐因素影响做题数据。

知识熟练度诊断任务

如前所述, 在图 1 模型框架中, U_u 是学生 u 的知识熟练度, 向量的每个值代表了学生 u 在不同知识点的掌握程度。为了学生的诊断结果是否精确, 引入了符合任务的新指标 DOA, 如公式(24)所示。从直觉上看, 如果学生 a 比学生 b 更好地掌握知识点 k , 那么 a 比 b 更有可能正确回答与 k 相关的题目。因此本实验采用一致度 (DOA) [26] 指标来评价学生在知识熟练度排序情况。

$$DOA(k) = \frac{1}{Z} \sum_{a=1}^N \sum_{b=1}^N \delta(F_{ak}^s, F_{bk}^s) \sum_{j=1}^M I_{jk} \frac{J(j, a, b) \wedge \delta(r_{aj}, r_{bj})}{J(j, a, b)} \quad (24)$$

其中 $Z = \sum_{a=1}^N \sum_{b=1}^N \delta(F_{ak}^s, F_{bk}^s)$, F_{ak}^s 是学生 a 在知识点 k 上的熟练度。 δ , I , J 均为指示函数, 如果 $x > y$, $\delta(x, y) = 1$, 否则 $\delta(x, y) = 0$ 。如果题目 j 考察了知识点 k , $I_{jk} = 1$, 否则 $I_{jk} = 0$ 。如果学生 a 和学生 b 均做了题目 j , 那么 $J(j, a, b) = 1$, 否则 $J(j, a, b) = 0$ 。也就是说, 在考虑学生题目和知识点两两关系之后, $DOA(k)$ 可以评价在 k 知识点上所有学生知识熟练度排序情况。那么, 对所有知识点的 $DOA(k)$ 进行求和取平均得到 DOA , 就可以评估诊断结果的质量 (即模型知识熟练程度诊断程度)。

在经典模型中, 我们只与 MIRT, PMF, QPMF 比较。因为这些模型考虑多值的学生特征向量, 其中 QPMF 还关联了知识点和题目之间的关系。基于知识熟练度诊断任务的特点, 本实验通过实验指标 DOA 来比较多数据集的诊断精度。图 5 为实验结果, 根据实验结果表有几点结论: 首先, BloomCDM 和 BloomCDM_RC 的 DOA 明显高于基线模型, 这证明了它们诊断出的知识熟练度是合理的且认知层级有利于校准知识熟练度的计算, 其次, BloomCDM_RC 不如 BloomCDM 精准, BloomCDM 存在一定的提升效果。此外, QPMF 比 PMF 的诊断效果要好, 说明 Q 矩阵信息对于获得可解释的诊断结果(知识熟练度向量)都很重要。

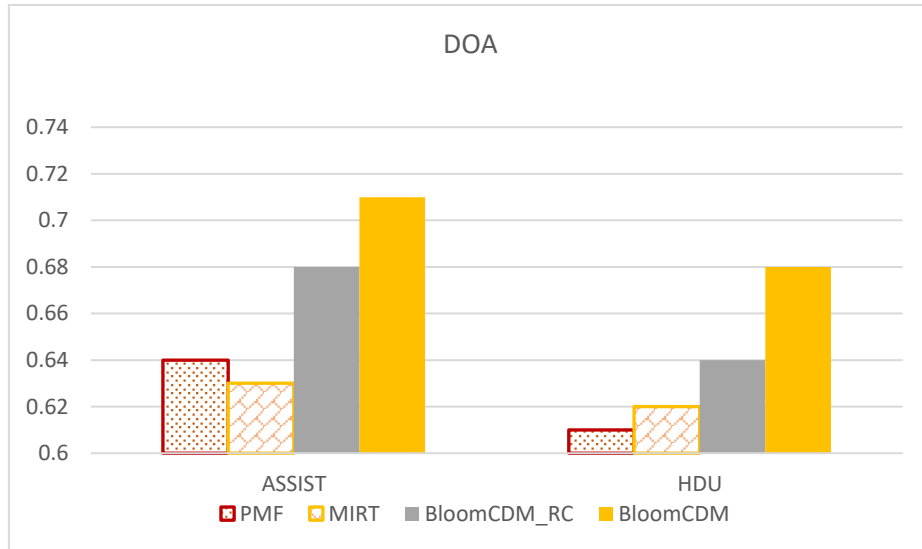


图 4 所有模型的知识熟练度任务性能

Fig. 4 Knowledge proficiency diagnosis performance for all knowledge points

可视化知识熟练度

在这里, 给出了一个学生知识熟练程度的认知诊断的例子。如图 3.13 所示, 雷达图显示了学生分别在 IRT, MIRT, PMF 和 BloomCDM4 种模型的知识熟练程度诊断结果。由于 IRT 只诊断了学生的一维能力, 因此能力在所有问题上都有相同的值且 IRT 的诊断结果在图 3.13 中是一个正多边形。因此, 与 IRT 相比, PMF 和 MIRT 可以提供更准确的知识诊断结果, 因为它们都诊断了学生的多维知识熟练度。但是 IRT 在第一个问题上得到了错误的结果, MIRT 和 PMF 诊断结果类似, 但在第三个问题上预测错误。只有 BloomCDM 对所有的三个问题预测都正确。



图 5 学生对知识点和三个问题的熟练程度的可视化
Fig. 5 Visualization of a student's proficiency on knowledge concepts and the parameters of three questions

8 总结与未来工作

针对以往知识熟练度计算模型的缺陷，即忽略了教育心理学中认知具有层级特点，本文提出了一种基于Bloom认知理论的知识熟练度计算模型BloomCDM。该模型的输入是学生在试题上的答题情况和由教育领域内的专家所标注的试题知识点关联矩阵(Q矩阵)。该模型的第一步是进行层级和层间的理论假设，抽象出数学模型并将题目特征向量投影到“知道”“领会”“应用”三层认知空间，

- (1) 知道建模：根据知道的理论定义进行假设，本文通过偏序 \succ_d 表示，从学生做题数据中学习题目的知道特征。
 - (2) 领会建模：根据领会的理论定义进行假设，领会层级以考察同一知识点的题目作为研究对象。假设相同知识点都被考察时，更进一步测量该知识点的理解度。本文通过**知识组**的形式构建属于该知识点的理解计算模型，捕获题目之间相同知识相关的特征并学习题目的领会特征。
 - (3) 应用建模：根据应用的理论定义进行假设，应用层级以考察相似知识点的题目作为研究对象。假设相似知识点都被考察时，更容易发现知识之间的内在联系。通过**高阶知识组**的形式构建跨知识点的计算模型，学习题目的应用特征。
 - (4) 为了解决从稀疏数据中获取层级结构信息的问题，基于本文的应用场景，设计了一个高阶知识组发现方法，可以从原始做题矩阵中发现高阶知识组结构并挖掘高阶知识组的结构信息，为基于 Bloom 认知理论的知识熟练度计算提供了良好的支撑。
- 最后，大量的实验证明了该模型可以精准地诊断学生的知识熟练度，证明了在知识熟练度计算中考虑 Bloom 认知层级是有益的，本文的设计是科学有效的。

在未来的研究中,我们将针对以下方面进行探索：

- (1) 本文采用概率图模型设计的BloomCDM的效果是显而易见的，但由于神经网络在非线性问题和特征嵌入方面表现出良好的性能^{错误!未找到引用源。}，可以轻松的达到概率模型达不到的准确度。是否可以构建可解释性的深度计算模型，在得到学生知识熟练度的基础上，有更好的学生表现预测。
- (2) 传统认知诊断模型MIRT在AUC指标上效果较好，甚至比推荐模型BPR，PMF效果要好，这是个很有趣的现象。本文认为随后的模型的研究点可以是在MIRT模型上添加知识点，来计算更为精确的知识熟练度。

Reference:

[3] De La Torre J. DINA model and parameter estimation: A didactic[J]. Journal of educational and behavioral statistics, 2009, 34(1): 115-130.

[4] QPMFChen Y, Liu Q, Huang Z, et al. Tracking knowledge proficiency of students with educational priors[C] Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017: 989-998.

[5] Piech C, Spencer J, Huang J, et al. Deep knowledge tracing[J]. arXiv preprint arXiv:1506.05908, 2015.

[6] Bloom B S. Taxonomy of educational objectives. Vol. 1: Cognitive domain[J]. New York: McKay, 1956, 20: 24.

[7] Harvey R J, Hammer A L. Item response theory[J]. The Counseling Psychologist, 1999, 27(3): 353-383.

[8] Reckase M D. Multidimensional item response theory models[M] Multidimensional item response theory. Springer, New York, NY, 2009: 79-112.

- [9] Sheng Y, Wikle C K. Bayesian multidimensional IRT models with a hierarchical structure[J]. Educational and psychological measurement, 2008, 68(3): 413-430.
- [10] Rost J. Rasch models in latent classes: An integration of two approaches to item analysis[J]. Applied Psychological Measurement, 1990, 14(3): 271-282.
- [11] Ma W, de la Torre J. A sequential cognitive diagnosis model for polytomous responses[J]. British Journal of Mathematical and Statistical Psychology, 2016, 69(3): 253-275.
- [12] Chen J, de la Torre J. A general cognitive diagnosis model for expert-defined polytomous attributes[J]. Applied Psychological Measurement, 2013, 37(6): 419-437.
- [13] Xiong L, Chen X, Huang T K, et al. Temporal collaborative filtering with bayesian probabilistic tensor factorization[C]. Proceedings of the 2010 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, 2010: 211-222.
- [14] Nedungadi P, Smruthy T K. Personalized multi-relational matrix factorization model for predicting student performance[M]. Intelligent Systems Technologies and Applications. Springer, Cham, 2016: 163-172.
- [15] Huang Z, Liu Q, Chen Y, et al. Learning or forgetting? A dynamic approach for tracking the knowledge proficiency of students[J]. ACM Transactions on Information Systems (TOIS), 2020, 38(2): 1-33.
- [16] 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.
- [17] 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, 34(04): 6153-6161.
- [18] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization[J]. Advances in neural information processing systems, 2007, 20: 1257-1264.
- [19] Moon T K. The expectation-maximization algorithm[J]. IEEE Signal processing magazine, 1996, 13(6): 47-60.
- [20] Carlin B P, Chib S. Bayesian model choice via Markov chain Monte Carlo methods[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1995, 57(3): 473-484.
- [21] White H. Maximum likelihood estimation of misspecified models[J]. Econometrica: Journal of the econometric society, 1982: 1-25.
- [22] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[J]. arXiv preprint arXiv:1205.2618, 2012.
- [23] Bottou L. Stochastic gradient descent tricks[M]. Neural networks: Tricks of the trade, Springer, Berlin, Heidelberg, 2012: 421-436.
- [24] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008: 9(11).
- [25] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. Kdd, 1996, 96(34): 226-231.
- [26] Huang Z, Liu Q, Chen E, et al. Question Difficulty Prediction for READING Problems in Standard Tests[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1).

附中文参考文献:

- [1] 余胜泉,汪晓凤. “互联网+” 时代的教育供给转型与变革[J].开放教育研究,2017,23(01):29-36.
- [2] 周庆,牟超,杨丹.教育数据挖掘研究进展综述[J].软件学报,2015,26(11):3026-3042.