



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于认知诊断的个性化习题推荐
作者：韩祎珂，徐彬，张硕
收稿日期：2023-09-05
网络首发日期：2023-12-19
引用格式：韩祎珂，徐彬，张硕. 基于认知诊断的个性化习题推荐[J/OL]. 计算机应用.
<https://link.cnki.net/urlid/51.1307.tp.20231218.1448.006>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于认知诊断的个性化习题推荐

韩伟珂¹, 徐彬^{1*}, 张硕²

(1. 东北大学 计算机科学与工程学院, 沈阳 110169; 2. 东北大学 信息科学与工程学院, 沈阳 110169)

(* 通信作者电子邮箱 xubin@mail.neu.edu.cn)

摘要:针对现有的基于认知诊断的习题推荐建模角度单一以及习题推荐结果不够合理的问题,提出了结合认知诊断和深度因子分解机的个性化习题推荐方法。首先,设计了一种全新的知识点关系计算方法,构建了课程知识树,并提出了增强 Q 矩阵的概念,能够准确表示习题所含知识点的关系。其次,提出了NeuralCD-KD模型,该模型计算了增强 Q 矩阵,利用特征二阶交叉和注意力机制融合习题难度内外因素,模拟学生的认知状态。通过在私有数据集和公开数据集验证了提出的认知诊断模型的有效性,并能够对学生的认知状态做出合理解释。为了个性化习题推荐,结合了诊断模型和深度双线性因子分解机,提出了NKD-DBFM方法,在私有数据集上验证了提出的习题推荐方法的有效性,在曲线下面积(AUC)指标下相较于最优基线模型NeuralCDM提升了3.7个百分点。

关键词:认知诊断;习题推荐;知识点关系; Q 矩阵;习题表征

中图分类号:TP391 **文献标志码:**A

Personalized exercise recommendation based on cognitive diagnosis

HAN Yike¹, XU Bin^{1*}, ZHANG Shuo²

(1. College of Computer Science and Technology, Northeastern University, Shenyang 110169, China;

2. College of Information Science and Technology, Northeastern University, Shenyang 110169, China)

Abstract: A personalized exercise recommendation method that combines cognitive diagnosis and deep factorization machine was proposed to address the problems of single modeling angle and unreasonable exercise recommendation results. Firstly, a new method for calculating the relationship between knowledge points was designed to construct a course knowledge tree, and the concept of enhanced Q matrix was proposed to accurately represent the relationship between knowledge points contained in exercises. Secondly, the NeuralCD-KD model was proposed to calculate the enhanced Q matrix, which uses second-order cross features and attention mechanism to fuse internal and external factors of exercise difficulty, and simulate students' cognitive status. The effectiveness of the proposed cognitive diagnosis model was verified on private and public datasets, and it can make reasonable explanations for students' cognitive status. To personalize exercise recommendation, the NKD-DBFM method was proposed by combining the diagnostic model with deep bilinear factorization machine, and the effectiveness of the proposed exercise recommendation method was verified on private datasets. Compared with the optimal baseline model NeuralCDM, the proposed method improved the area under the curve (AUC) index by 3.7 percentage points.

Key words: cognitive diagnosis; exercise recommendation; knowledge point relationship; Q matrix; exercise representation

0 引言

在当今信息技术高速发展的背景下,海量教学资源涌入给学生带来了困扰。推荐技术成为解决这一难题的有力工具。利用认知诊断(Cognitive Diagnosis, CD)模型从知识点层面刻画学生的知识状态,再针对学生个性化认知状态进行习题推荐已经成为当下研究者们关注的一个热点。本研究基于我校C语言程序设计课程以及另一开放数据集,结合知识点关系和习题表征,提出了增强 Q 矩阵的概念,提出基于知识点关系和习题表征的认知诊断(Neural Cognitive Diagnosis with Knowledge-based Discernment, NeuralCD-KD)模型结合

增强 Q 矩阵用于更真实地建模学生的知识状态,在此基础上设计出结合认知诊断和深度因子分解机(Neural Knowledge-based Cognitive Diagnosis with Deep Bilinear Factorization Machines, NKD-DBFM)的习题推荐方法实现习题推荐。

1 相关工作

现有的习题推荐方法可分为四类^[1]:基于知识的习题推荐、基于协同过滤(Collaborative Filtering, CF)的习题推荐^[2]、基于认知诊断的习题推荐和基于深度知识追踪的习题推荐。CD模型中的 Q 矩阵定义了习题和知识点之间的关联,在习题推荐领域,目前最常应用的认

收稿日期:2023-09-05;**修回日期:**2023-12-03;**录用日期:**2023-12-05。 **基金项目:**辽宁省自然科学基金面上项目(2022-MS-119);中国高校产学研创新基金-云中大学项目(2022MU017);国家自然科学基金面上项目(72271048)。

作者简介:韩伟珂(1999—),女,河南新密人,硕士研究生,主要研究方向:故障诊断、智慧教育;徐彬(1980—),男,浙江兰溪人,副教授,博士,CCF会员,主要研究方向:人工智能、智慧教育、工业智能、故障诊断;张硕(1988—),男,辽宁沈阳人,博士研究生,主要研究方向:异常检测、机器学习、需求预测。

知诊断模型包括项目反应理论(Item Response Theory, IRT)^[3]和DINA (Deterministic Input, Noisy AND-Gate)^[4]模型。除此之外,还有一些相关研究提出了不同的方法和模型来改进认知诊断和习题推荐的效果。SU等人^[5]针对现有模型只考虑学生与问题(或学生与技能)之间的双向交互,忽略了实体之间的潜在高阶关系的问题,提出了一种基于图的认知诊断模型(Graph-based cognitive diagnosis Module, GCDM)。该模型通过异构认知图直接发现学生、技能和问题之间的交互关系。WANG等人^[6]设计了神经认知诊断模型(Neural Cognitive Diagnosis Model, NeuralCDM),利用神经网络对学生的知识状态进行诊断。LI等人^[7]首先利用神经网络获取练习的多维特征(如难度、区分度、猜测率和失误率),然后用项目反应理论和神经网络来描述学生和练习之间的交互,以确定学生的认知状态,同时引入单调性假设和数据预处理机制到神经网络中,提高了诊断结果的准确性。张所娟等人^[8]提出了一种融合知识交互关系的认知诊断深度模型,该模型实现了学习者认知状态和知识权重的统一表示,同时提出了基于Choquet积分的理想作答反应计算算法,并通过模糊测度的深度神经网络来预测学生的答题情况。汤成^[9]考虑到DINA模型无法完全包含潜在因素,引入遗忘因素和知识点作答次数等因素,提出了一种基于期望最大化算法的模型。该模型通过计算学生的知识掌握状态,利用知识状态、学生得分情况和 Q 矩阵进行分解,并通过学生对知识点的掌握率来进行推荐。朱天宇^[10]等的研究提出了一种基于学生知识点掌握程度和协同过滤的个性化试题推荐方法,通过综合考虑学生的学习状态和相似学生的共性,该方法在保持试题推荐解释性的同时提高了可靠性。

此外,推荐领域常用的模型包括因子分解机(Factorization Machines, FM)^[11]、域感知分解机(Field-aware Factorization Machines, FFM)^[12]以及深度因子分解机(Deep Factorization Machine, DeepFM)^[13]等,它们可用于习题推荐中的特征表示学习,从而更好地捕捉习题和学生之间的复杂关系,提高推荐准确性和效果。DeepFM通过结合FM和深度神经网络(Deep Neural Networks, DNN)^[14],不仅能学习特征间的一阶和二阶交互,还能发掘隐含的高阶信息,实现低阶和高阶特征交叉。

2 NKD-DBFM 习题推荐方法

NKD-DBFM 习题推荐方法的整体结构如图1所示。整个过程可以大致分为三个阶段:数据准备、学生认知建模和习题推荐。

2.1 知识点及习题因素的表征

数据准备阶段主要是获得知识点及习题因素的表征,需要得到知识点、 Q 矩阵、课程知识树、习题难度特征集以及对习题难度进行标注。

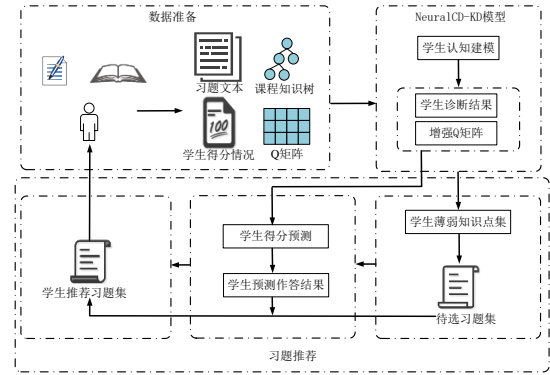


图1 NKD-DBFM 的整体结构

Fig. 1 The overall architecture of NKD-DBFM

认知诊断中虽然 Q 矩阵通常用0和1表示习题是否涉及某些知识点,但实际上习题考察知识点的比例也会影响学生的答题结果。通过课程知识树,可以建模知识点之间的结构关系以及习题文本和 Q 矩阵中知识点的共现关系,从而实现对习题考察知识点比例的建模即得到增强 Q 矩阵。在这之前,可以通过Jieba分词对教材文本进行分词,获取候选词,并通过词频-逆文档频率^[15](Term Frequency-Inverse Document Frequency, TF-IDF)方法获得权重较高的有用词。然后通过人工筛选,得到课程知识点。获得知识点后,人工标注习题涉及的知识点,从而得到 Q 矩阵,并结合教师意见构建课程树。关于如何获得习题考察知识点比例的建模,在学生认知建模中进行介绍。

在IRT模型和DINA模型中,习题难度起到了重要作用。基于NeuralCDM模型的基线方法中,使用了知识点难度和习题区分度,但是只使用习题的ID作为输入,并不合理。传统模型只利用题目的正确提交率来扩展习题难度,但习题的正确率会受到提交量的影响,因此需要更准确的方法来建模习题难度。

通过答题记录,可以构建习题难度特征集,例如习题的平均有效答题时间,时间越长表示难度越大。通过人工提取这些特征,并使用Feature-Selector工具去除高缺失值和高相关性特征,即可获得最终的习题难度特征集。以私有数据集为例,最终的习题难度特征包括: avg_time(平均有效作答时间)、std_time(有效作答习题的时间方差)、code_length(平均正确作答代码长度)、r_rate(正确作答次数与该题总提交次数之比)、one_rate(一次作答成功次数与总提交次数之比)、sim_rate(正确提交中标记为高度相似的次数与总正确提交次数之比)。获得难度特征集后,可以通过Kmeans算法对习题难度进行聚类分级,最后将习题难度分为三个等级。

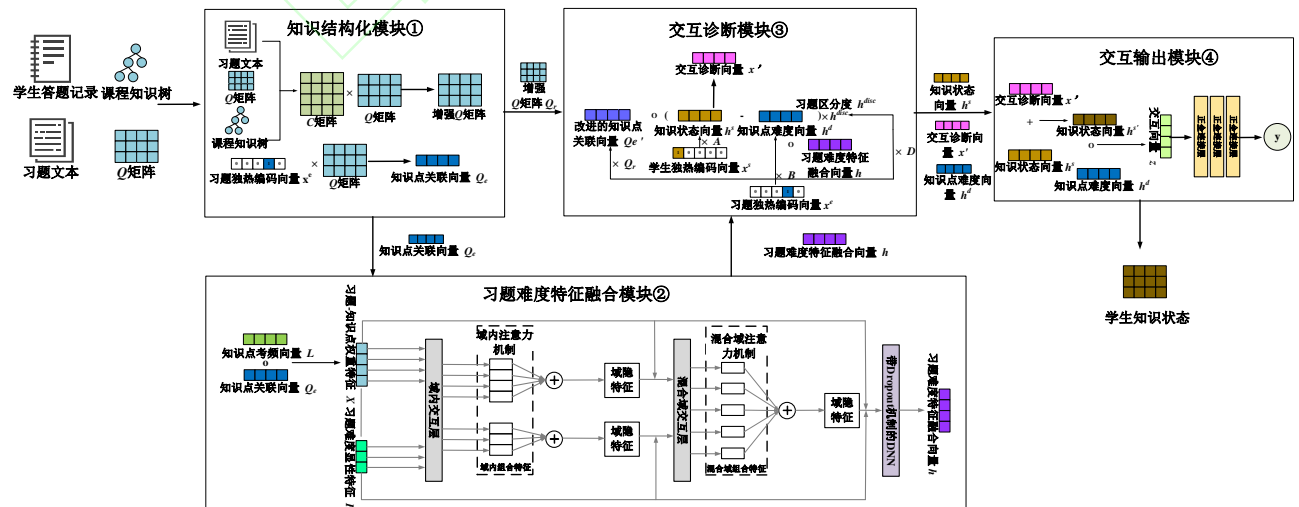


图2 NeuralCD-KD 模型整体框架

Fig. 2 The overall framework of NeuralCD-KD Model

2.2 学生认知建模

用于学生认知建模的 NeuralCD-KD 模型整体框架如图 2 所示。该模型由四个模块组成:知识结构化模块、习题难度特征融合模块、交互诊断模块和交互输出模块。在知识结构化模块中,利用课程知识树明确知识点之间的结构关系,并结合 Q 矩阵和题目文本来表达知识点之间的关系,从而获得增强的 Q 矩阵。通过习题难度特征融合模块,将习题的知识点权重和习题难度的显性特征计算得到习题难度特征融合向量。然后,将增强的 Q 矩阵和习题难度特征融合向量输入到交互诊断模块中,得到交互诊断向量。最后,将得到的交互诊断向量、学生知识状态向量和知识点难度向量送入交互输出模块,用于预测学生的答题表现。训练得到的知识状态向量则作为学生认知诊断的结果。

2.2.1 知识化结构模块

为了改进 NeuralCDM 中丢失知识关联的问题,知识结构化模块利用课程知识树、 Q 矩阵和习题文本构建了知识点关系矩阵 C ,同时利用习题独热编码和 Q 矩阵获得知识点关联向量 Q_e 。构建矩阵 C 涉及三个不同来源的知识点关系矩阵: C_1 代表从课程知识树计算得到的知识点结构关系矩阵, C_2 代表结合 Q 矩阵和题目文本关键词的共现矩阵, C_3 代表通过直接计算 Q 矩阵的共现率得到的矩阵。矩阵 C 与 Q 矩阵相乘后归一化就得到了增强 Q 矩阵,能够准确展现习题对知识点的考察频率。计算矩阵 C 的方法如下:

$$C = \lambda C_1 + (1 - \lambda)(C_2 + C_3) \quad (1)$$

其中, λ 是比例因子,大小在 0 到 1 之间。 C_1 的计算是从课程知识树出发的,先要对课程知识树先序遍历,得到每个节点的编号。在该树中,任意两个知识点都会有唯一的路径相连。本文可以将知识点 k_i 与 k_j 之间的相连路径定义为 G , G 的定义如下:

$$G = \{(k_i, v_1, k_p), (k_p, v_2, k_q), \dots, (k_{j-1}, v_g, k_j)\} = \{g_1, g_2, \dots, g_g\} \quad (2)$$

其中, v 是两个节点之间的边,下标 g 由路径中经过的边数决定,在 $g_i = (k_i, v_i, k_p)$ 中,将边分为两类, v^+ 和 v^- 。若 $i > p$, $v_i \in v^+$, $i < p$, $v_i \in v^-$ 。知识点 k_i 和 k_j 的结构关系 c_{ij} 的计算如下:

$$c_{ij} = \begin{cases} \frac{1}{2^{v^+ - v^-} \times \text{depth}(k_j)} & i > j \\ 1 & i = j \\ 0 & i < j \end{cases} \quad (3)$$

其中, V^+ 和 V^- 分别代表 v^+ 和 v^- 中边的个数, $\text{depth}(k_j)$ 表示的是知识点 k_j 处于树中的深度。假设课程涉及 K 个知识点,那么 $C_1 \in R^{K \times K}$ 。当 $i < j$ 时,即知识点 k_i 在被学习时的先后顺序要先于 k_j ,此时在学习 k_i 时并不需要依赖 k_j 的知识作为铺垫。因此,在这种情况下, $c_{ij} = 0$ 。 c_{ij} 的值越大表示两个知识点之间的联系越紧密,否则关联性就越弱。

共现率的计算公式如下:

$$R(t_1, t_2) = \frac{P(t_1, t_2)(f(t_1) + f(t_2))}{f(t_1) \times f(t_2)} \quad (4)$$

其中, $P(t_1, t_2)$ 表示 t_1 和 t_2 出现在同一个题目中的次数, $f(t_1)$, $f(t_2)$ 表示 t_1 , t_2 在题目中出现的总次数。在计算 C_2 时, t_1 , t_2 代表关键词,计算 C_3 时,则代表知识点。此外由于关键词总量比知识点总量更大,会造成矩阵稀疏,因此计算 C_2 时需将关键词聚合,使得多个关键词指向同一个知识点得到 C_2 , C_3 则直接计算 Q 矩阵中知识点的共现率得到。

2.2.2 习题特征融合模块

该模块通过综合习题难度的内外因素解决主习题难度表征不准确的问题。其中,题目的正确提交率、平均作答时间等是习题难度的具体外在表现,而习题考察的知识点频率也是影响题目难度的重要因素,被视为题目难度的内在体现。知识点的难度与权重成正相关^[16]。

2.1 小节构造的习题难度特征集与学生答题记录合并得到新的答题数据,新数据采用随机森林进行特征选择,最后选出最能影响学

生答题情况的前 N 个特征即为单个习题的难度显性特征 $F = (f_1, f_2, \dots, f_N)$, F 为 N 维行向量。计算知识点的考察频率,将知识点的考察频率组成的向量 L 与知识点关联向量 Q_e 按元素相乘即可得到习题难度的内在因素 $X = (x_1, x_2, \dots, x_K)$, X 是 K 维行向量。 X 与 F 拼接后就得到了输入 $U \in R^{1 \times (N+K)}$ 。

将特征进行组合相对于单独使用特征更有利于模型预测^[17],而特征的二阶交叉是推荐领域常用的组合特征的方法,考虑到不同的组合特征项的重要程度存在差异,引入注意力机制^[18]实现域内和混合域组合特征项重要性的计算。域交互层实现特征组合,域隐特征的计算过程相似,具体计算如下:

$$e_{ij} = \text{ReLU}(W(w_{i,j}x_i x_j) + b) \quad (5)$$

$$a_{ij} = \frac{\exp(e_{ij}h_{e_{ij}})}{\sum_{x_{i,j}} \exp(e_{ij}h_{e_{ij}})} \quad (6)$$

$$\beta = \sum_{x_{i,j}} a_{ij} \cdot w_{i,j}x_i x_j \quad (7)$$

其中: $W, w_{i,j}, b, h_{e_{ij}}$ 为可学习参数, $w_{i,j}x_i x_j$ 是特征组合项, a_{ij} 为注意力分数, $x_{i,j}$ 所属的域不同,可学习参数的值不同,计算得到的域隐特征也不同, $x_{i,j} \in F$, 得到习题知识点权重域隐特征 β_1 , $x_{i,j} \in X$, 得到习题难度显性特征域隐特征 β_2 , $x_{i,j} \in \{U, \beta_1, \beta_2\}$, 得到混合域隐特征 β_3 , 将 $\{U, \beta_3\}$ 送入 DNN 实现的融合输入层,得到最终的习题难度特征融合向量 $h \in R^{1 \times K}$ 。

2.2.3 交互诊断模块

该模块与原模型计算相似,改变在于增强 Q 矩阵、习题难度特征融合向量 h 参与运算。模块具体计算如下:

$$Q'_e = x^e \times Q_r \quad (8)$$

$$h^s = \text{sigmoid}(x^s \times A) \quad (9)$$

$$h^d = \text{sigmoid}(x^e \times B \circ h) \quad (10)$$

$$h^{disc} = \text{sigmoid}(x^e \times D) \quad (11)$$

$$x' = Q'_e \circ (h^s - h^d) \times h^{disc} \quad (12)$$

其中, A, B, D 为可训练矩阵,通过增强 Q 矩阵获得改进的知识点关联向量 Q'_e ,通过习题难度特征融合向量得到知识点难度向量 h^d ,最后通过综合两者以及学生认知状态 h^s 、习题区分度 h^{disc} 得到学生诊断因素 x' 。

2.2.4 交互输出模块

交互输出模块在于模拟学生答题过程中知识状态的进化,不考虑抄袭、猜测的前提下,无论学生是否答对题目,在思考后,会加深对知识点的理解。将学生未答题时的知识状态向量 h^s 与交互诊断向量 x' 结合,得到学生的潜在知识状态即答题后的状态知识向量 h'' ,潜在知识状态与习题的难度向量 h^d 结合得到交互向量,并且通过全连接层来学习隐藏信息,最终得到学生对题目的作答结果预测值,模型使用交叉熵作为损失函数。

2.3 习题推荐

2.3.1 生成待选习题集

将根据 NeuralCD-KD 模型得到的学生潜在知识状态 h'' 作为最终的学生对各知识点的掌握状态向量,对应的值越高代表学生对该知识点的掌握程度越好,将状态值递增排序后,将前 $K/2$ 个值对应的知识点作为学生的薄弱知识点。

遍历学生薄弱知识点的答题记录,根据习题难度分类,依次计算困难、中等、简单习题学生的回答正确率,当正确率不低于 0.8 时考虑将难度高一类的习题加入待选习题集否则继续判断下一难度习题的正答率,即当中等习题正答率大于等于 0.8 时将困难习题加入待选习题集,否则继续判断简单习题的正答率,特别的当困难习题正答率较高时,将未作答的困难习题加入待选习题集,当简单习题的正答率都比较低时,只将简单习题加入待选习题集,最终得到对学生更有挑战

性的更有学习必要的待选习题集。

2.3.2 学生得分预测

DeepFM 通过两个嵌入向量的内积实现低阶交互,但是没有考虑到嵌入向量不同维度上的值的重要性不同,为了更好地捕捉特征间复杂关系,提出了深度双线性因子分解机(Deep Bilinear Factorization Machines, DeepBFM),通过在二阶交互部分对稀疏特征加入双线性交互特征来实现区分交互向量在不同维度上的重要性,模型结构如图3所示。

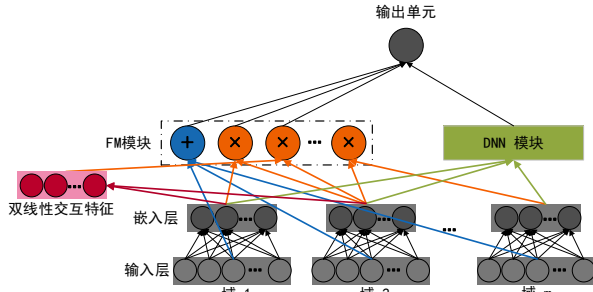


图3 DeepBFM 的模型结构

Fig. 3 Model structure of DeepBFM

假设稀疏特征嵌入后对应的矩阵 $E = [e_1, e_2, \dots, e_f]$, f 是稀疏特征域的数量,双线性交互特征的计算如下:

$$p_i = e_i \cdot W_{ij} \circ e_j \quad (13)$$

其中, $W_{ij} \in R^{k \times k}$, 是可学习参数矩阵, \circ 是哈达玛积, 进而得到特征矩阵 $P = [p_1, p_2, \dots, p_i, p_{i+1}, \dots, p_n] \in R^{n \times k}$, $n = \frac{f(f-1)}{2}$, k 为 DeepFM 中给定的特征向量对应的隐向量的维度。将 P 和 E 拼接后送入 FM 中参与交叉部分运算, DNN 输入不变, 计算过程同 DeepFM, 损失函数采用交叉熵损失函数。

实验时输入学生知识状态 h'' 、学生 ID、习题 ID 和增强 Q 矩阵即学生知识点交互域、学生域、习题域、习题知识点交互域四个域, 这样既实现了学生和习题的全局交互, 也实现了学生和习题在知识属性上的局部交互, 最终得到学生答对习题的预测概率向量。

2.3.3 确定推荐习题集

根据待选习题集以及预测的答对概率, 设置阈值 $[\lambda_1, \lambda_2]$ ($\lambda_1 < \lambda_2 < 1$), 筛选学生答对概率在 λ_1 和 λ_2 之间的待选习题集作为最终的推荐习题集, 阈值范围可以根据学习时期有所调整, 学习初期可选择较高的阈值, 学习后期可将阈值调低。

3 实验

3.1 实验数据

实验涉及两个数据集, NEUOJ 和 Assistments2009-2010 Skill-Builder(简记为 Assistments2009-2010), 具体数据集描述见表1。

NEUOJ 是来自东北大学在线编程社区的私有数据集。该数据集从 C 语言题库中选取了部分题目以及相关的答题数据, 数据具体包括题目文本、学生答题记录等。在实验中, 2.1 节中对习题表征进行处理后得到的习题难度特征集也作为部分原始数据。

Assistments2009-2010 是一个公开的数据集, 来源于在线辅导平台 ASSISTments。该数据集提供了学生的答题记录, 但没有公开题目的文本和知识结构关系。因此, NeuralCD-KD 模型并没有在该数据集上进行验证, 并且基于 NeuralCD-KD 模型的习题推荐方法 NKD—DBFM 也没有在该数据集上实现。

在对 NEUOJ 和 Assistments2009-2010 数据集进行实验之前, 进行了随机划分, 将 80% 的数据作为训练集, 剩余的数据作为测试集。

3.2 认知诊断实验

整个模型基于 Pytorch 实现, 采用 Adam 优化器消除其余因素的影响, 学习率为 0.002, dropout 概率设为 0.5, DNN 层数为 2, 神经元数

为 256 和 K, 激活函数为 ReLu, 交互输出模块的神经元数分别设为 512, 256, 并且习题难度显性特征数量经过实验后选用 4。

表1 数据集概况

Tab. 1 Overview of datasets

数据集	学生人数	习题数	知识点数	答题记录数
NEUOJ	6231	400	22	283125
Assistments2009-2010	4163	17746	123	324572

为了验证所提出模型的实际效果, 根据不同的数据集, 将 NeuralCD-KD 与其他多个 CD 模型进行对比并进行消融实验, CD 模型包括 DINA、IRT、MIRT 等经典 CD 模型, 以及本文的基础模型 NeuralCDM, 消融实验包括 N-K、N-D、N-KD 等方法。N-K 不考虑知识结构化模块的改进, 在习题难度特征融合模块只输入习题知识点权重特征; N-D 不考虑知识结构化模块的改进, 在习题难度特征融合模块只输入习题难度显性特征; N-KD: 不考虑知识结构化模块的改进, 在习题难度特征融合模块同时输入习题难度显性特征和习题知识点权重特征; 以上实验在交互诊断模块均使用知识点关联向量 Q_e , 而不是改进后的 Q_e' 。

评测指标选用 AUC (Area Under Curve)、均方根误差 (Root Mean Square Error, RMSE) 和预测准确率 (Accuracy, Acc)。实验结果见表 2, 从实验结果可以看出, 在 NEUOJ 数据集上, 相比于其他 CD 模型, NeuralCD-KD 在 Acc 上提升 0.1 到 8.1 个百分点, 在 RMSE 上下降 0.2 到 4.4 个百分点, 在 AUC 上提升 0.3 到 8 个百分点, 消融实验的效果也都优于其他 CD 模型, 因此设计的模型的各个改进之处都对模型效果的提升起到积极的作用, 并且 NeuralCD-KD 达到最好的认知诊断效果。在 Assistments2009-2010 数据集上, 选取习题难度显性特征是习题的正答率和平均用时, N-KD 方法比 NeuralCDM 在 Acc 和 AUC 上提升 0.5 个百分点, 在 RMSE 上降低 1.1 个百分点, 是消融实验中效果最好的方法, N-K 表现最差也有 0.3 个百分点的 AUC 的提升, 证明了所提改进方法不仅适用于私有数据集, 而且在公开数据集上也取得了效果提升。

表2 NeuralCD-KD 模型与对比模型评估指标对比

Tab. 2 Comparison of evaluation indexes between NeuralCD-KD model and comparison model

Model	NEUOJ			Assistments2009-2010		
	Acc	RMSE	AUC	Acc	RMSE	AUC
IRT	0.659	0.469	0.660	0.675	0.463	0.685
DINA	0.643	0.473	0.649	0.651	0.467	0.678
MIRT	0.680	0.464	0.684	0.698	0.462	0.717
NeuralCDM	0.717	0.439	0.718	0.725	0.438	0.750
N-K	0.720	0.435	0.723	0.726	0.430	0.753
N-D	0.722	0.433	0.724	0.728	0.432	0.754
N-KD	0.723	0.431	0.726	0.730	0.427	0.755
NeuralCD-KD	0.724	0.429	0.729			

学生对各个知识点的掌握程度是对学生进行认知诊断的最终结果, 以其中两名学生作答的四道题为例, 诊断结果展现了两者的知识状态的差异性。表 3 是展示四道习题考察知识点的 Q 矩阵以及两名学生的作答结果, 而表示习题对各知识点的考察比例的增强 Q 矩阵如表 4。

图 4 和图 5 分别是二人对应的诊断结果。通过柱形图和折线图分别展示学生的诊断结果和习题的难度分布以学生的诊断结果为例, 根据 Q 矩阵可知习题 1 的考察知识点为知识点 6, 其难度值为 0.46。根据增强 Q 矩阵可以发现, 习题 1 在知识点 6 和知识点 1 上的考察比例最大。学生甲对知识点 1 和 6 的掌握水平值均大于对应知识点的难度值, 因此能正确回答习题 1。虽然习题 1 在知识点 4 上的难度值为 0.85, 大于学生甲对知识点 4 的掌握水平值 0.69, 但是习题 1 实际考察知识点 4 的成分很少, 根据增强 Q 矩阵可知仅为 0.05, 因此并不影响学生的答题结果。

对比两人的诊断结果可以看出, 学生甲的知识掌握均衡, 对知识

点的掌握程度均高于0.5,而学生乙对知识点掌握呈现两极分布,部分知识点的掌握程度大于0.75,而部分知识点的掌握程度甚至低于0.35。对于知识点4,学生甲的知识状态值为0.69,学生乙的知识状态值为0.83,应建议学生甲加强对知识点4的学习。而学生乙对于知识点1、2和6的掌握程度均小于0.35,因此这三个知识点将会被列入学生的薄弱知识点集,参与后续的习题推荐。

表3 习题的Q矩阵及学生作答结果

Tab. 3 Q matrix of exercises and the results of students' answer

习题	知识点1	知识点2	知识点3	知识点4	知识点5	知识点6	甲	乙
习题1	0	0	0	0	0	1	■	■
习题2	0	1	0	0	0	0	■	■
习题3	0	0	1	1	1	0	■	■
习题4	1	1	0	0	0	0	■	■

表4 习题的增强Q矩阵

Tab. 4 Enhanced Q matrix of exercises

习题	知识点1	知识点2	知识点3	知识点4	知识点5	知识点6
习题1	0.33	0.12	0.05	0.03	0.05	0.42
习题2	0.05	0.42	0.12	0.10	0.08	0.23
习题3	0.01	0.08	0.27	0.22	0.30	0.12
习题4	0.37	0.35	0.11	0.14	0.01	0.02

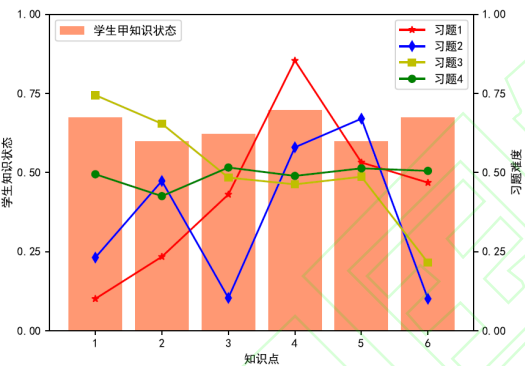


图4 学生甲的诊断结果

Fig. 4 Student A's diagnosis

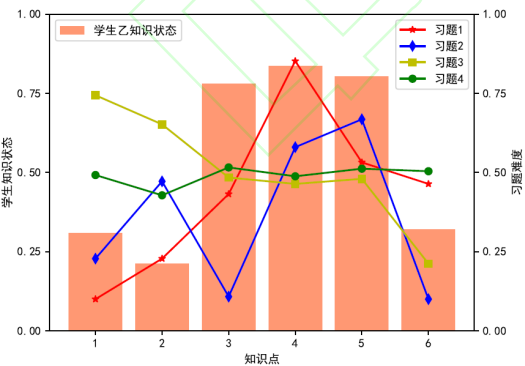


图5 学生乙的诊断结果

Fig. 5 Student B's diagnosis

3.3 习题推荐实验

学习率设定为0.001,采用Adam优化器排除无关因素的影响,DNN层数为3,对应神经元数目256,128,1,激活函数为Relu。

为了验证NKD-DBFM模型的有效性,在NEUOJ数据集上将其与几种方法做对比实验,包括NeuralCD-KD、FM、NKD-FM和NKD-DFM。FM方法以学生习题得分情况和Q矩阵作为输入,NKD-FM和NKD-DFM输入与NKD-DBFM相同,NKD-FM使用FM模型评估学生

答题结果而NKD-DFM使用DeepFM预测学生答题结果,所有对比方法都选用学生答对概率最高的N道习题作为最终的推荐习题。学生成绩预测选用AUC作为评价标准,并在困难等级习题上进行推荐精确率(Precision)的实验^[1],实验时为每个学生推荐5道习题,最后以作答正确率SR^[19]为评价标准进行习题难度设定实验。

学生成绩预测实验结果如表5第一列所示。本文提出的NKD-DBFM方法在AUC上相较于其他方法提升了0.2到2.6个百分点,并且根据上一小节的学生诊断实验可知,相对于最优的基线模型NeuralCDM,在AUC上提升了3.7个百分点,NKD-FM与FM相比,提升了1.1个百分点,了NKD-FM缓解了输入数据特征量的不足,而学生知识状态和增强Q矩阵含有更多的信息量能够提升模型效果;NKD-DBFM与NKD-FM和NKD-DFM的对比分别提升了1.1和0.2个百分点,说明将利用双线性操作改进后的低阶交互与高阶交互结合起来比起低阶交互和普通内积的二阶交互能达到更好的预测效果。

表5 NKD-DBFM与对比方法的成绩预测结果

Tab. 5 Performance prediction results for NKD-DBFM and comparison methods

方法	AUC	Precision
NeuralCD-KD	0.729	0.918
FM	0.733	0.926
NKD-FM	0.744	0.928
NKD-DFM	0.753	0.934
NKD-DBFM	0.755	0.936

习题推荐实验结果如表5第二列所示,NKD-DBFM方法在precision上提升0.2到1.8个百分点,精确率达到0.936,取得了最佳效果,说明该方法比其他方法推荐效果更优。

关于习题难度的设定实验如表6所示。图中阈值范围被分为5个区间,每个区间阈值变化量为0.2。随着阈值的增大,被推荐的习题难度变小,学生答对的概率也随之增加,这证明了本文的方法可以帮助学生推荐符合设定阈值范围的习题。

表6 不同阈值段下的SR对比

Tab. 6 Comparison of SR under different threshold segments

阈值范围	SR
[0,0.2)	0.18
[0.2,0.4)	0.39
[0.4,0.6)	0.56
[0.6,0.8)	0.76
[0.8,1.0)	0.90

4 结语

通过研究基于认知诊断的个性化习题推荐方法,本文改进了现有的推荐算法并提出了NKD-DBFM方法。实验结果显示,该方法在预测学生作答情况、推荐习题的精确率等方面取得了良好的效果和可解释性。未来的研究可以在Q矩阵自动标注和学科领域迁移等方面进行改进,进一步提高方法的性能和适用性。

参考文献 (References)

[1] 周安伟. 基于神经认知诊断的个性化试题推荐方法研究[D]. 兰州: 西北师范大学, 2021: 2-5. DOI: 10.27410/d.cnki.gxbfu.2021.001749. (ZHOU A W. Neural cognitive diagnosis model recommendation based on research on personalized test items [D]. Lanzhou: Northwest Normal University, 2021: 2-5. DOI: 10.27410/d.cnki.gxbfu.2021.001749.)

[2] HERLOCKER J L, KONSTAN J A, BORCHERS A, et al. An algorithmic framework for performing collaborative filtering[C/OL]// Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. New York: ACM, 1999: 230-237 [2023-06-27]. https://dl.acm.

- org/doi/10.1145/312624.312682. DOI:10.1145/312624.312682.
- [3] TAY L, MEADE A W, CAO M. An overview and practical guide to IRT measurement equivalence analysis [J/OL]. *Organizational Research Methods*, 2015, 18 (1): 3-46. DOI: 10.1177/1094428114553062.
- [4] DE LA TORRE J. DINA model and parameter estimation: a didactic [J]. *Journal of Educational and Behavioral Statistics*, 2009, 34(1): 115-130.
- [5] SU Y, CHENG Z, WU J, et al. Graph-based cognitive diagnosis for intelligent tutoring systems [J/OL]. *Knowledge-Based Systems*, 2022, 253: 109547. DOI: 10.1016/j.knosys.2022.109547.
- [6] WANG F, LIU Q, CHEN E, et al. Neural cognitive diagnosis for intelligent education systems [M/OL]. arXiv, 2020 [2023-06-27]. <http://arxiv.org/abs/1908.08733>. DOI: 10.48550/arXiv.1908.08733.
- [7] LI G, HU Y, SHUAI J, et al. NeuralNCD: A neural network cognitive diagnosis model based on multi-dimensional features [J/OL]. *Applied Sciences*, 2022, 12 (19): 9806. DOI: 10.3390/app12199806.
- [8] 张所娟,余晓晗,陈恩红等. 融合知识交互关系的认知诊断深度模型[J]. *模式识别与人工智能*, 2023, 36(01): 22-33. DOI: 10.16451/j.cnki.issn1003-6059.202301002. (ZHANG S J, YU X H, CHEN E H, et al. A concept interaction-based cognitive diagnosis deep model [J]. *Pattern Recognition and Artificial Intelligence*, 2023, 36 (01): 22-33. DOI: 10.16451/j.cnki.issn1003-6059.202301002.)
- [9] 汤成. 基于深度学习与认知诊断的教学资源推荐算法研究[D]. 广州: 华南理工大学, 2020: 1. DOI: 10.27151/d.cnki.ghnl.2020.000492. (TAN G C. Research on educational resource recommendation algorithm based on deep learning and cognitive diagnosis [D]. Guangzhou: South China University of Technology, 2020: 1. DOI: 10.27151/d.cnki.ghnl.2020.000492.)
- [10] 朱天宇,黄振亚,陈恩红等. 基于认知诊断的个性化试题推荐方法[J]. *计算机学报*, 2017, 40(1): 16. (ZHU T Y, HUANG Z Y, CHEN E H, et al. Cognitive Diagnosis Based Personalized Question Recommendation [J]. *Chinese Journal of Computers*, 2017, 40(1): 16.)
- [11] RENDLE S, 2010. Factorization machines [C/OL]//2010 IEEE International Conference on Data Mining. 995-1000. DOI: 10.1109/ICDM.2010.127.
- [12] JUAN Y, ZHUANG Y, CHIN W S, et al. Field-aware factorization machines for CTR prediction [C/OL]//Proceedings of the 10th ACM Conference on Recommender Systems. New York, NY, USA: Association for Computing Machinery, 2016: 43-50 [2023-06-26]. <https://doi.org/10.1145/2959100.2959134>. DOI:10.1145/2959100.2959134.
- [13] GUO H, TANG R, YE Y, et al. DeepFM: A factorization-machine based neural network for CTR prediction [M/OL]. arXiv, (2017-03-13) [2023-06-27]. <http://arxiv.org/abs/1703.04247>. DOI:10.48550/arXiv.1703.04247.
- [14] 孙志军,薛磊,许阳明等. 深度学习研究综述[J]. *计算机应用研究*, 2012, 29(08): 2806-2810. (SU N Z J, XUE L, XU Y M, et al. Overview of deep learning [J]. *Application Research of Computer*, 2012, 29(08): 2806-2810.)
- [15] YU Z, ZHUO M L, FU J R, et al. Semi-automatic Emotion Recognition from Textual Input Based on the Constructed Emotion Thesaurus [J/OL]. 2005 International Conference on Natural Language Processing and Knowledge Engineering, 2005: 571-576. DOI:10.1109/NLPKE.2005.1598802.
- [16] CHEN J, ZHANG H, HE X, et al. Attentive collaborative filtering: multimedia recommendation with item-and component-level attention [J/OL]. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017: 335-344. DOI:10.1145/3077136.3080797.
- [17] 李忧喜,文益民,易新河等. 一种改进的模糊认知诊断模型[J]. *数据采集与处理*, 2017, 32 (05): 958-969. DOI: 10.16337/j.1004-9037.2017.05.013. (LI Y X, WEN Y M, YI X H, et al. Revised model of fuzzy cognitive diagnosis framework [J]. *Journal of Data Acquisition & Processing*, 2017, 32 (05): 958-969. DOI: 10.16337/j.1004-9037.2017.05.013.)
- [18] YU H F, LO H Y, HSIEH H P, et al. Feature engineering and classifier ensemble for KDD cup 2010 [C/OL]//Knowledge Discovery and Data Mining. 2010 [2023-06-27]. <https://www.semanticscholar.org/paper/Feature-Engineering-and-Classifer-Ensemble-for-KDD-Yu-Lo/35793d9bed510d51cb45c54a4ba36132d4929fb9>.
- [19] 李猛. 基于神经认知诊断的学习分析与个性化习题推荐研究[D]. 江西: 江西师范大学, 2021: 34-35. DOI: 10.27178/d.cnki.gjxsu.2021.000455. (LI M. Cognitive diagnosis exercise recommendation method based on neural research on learning analysis and personalized [D]. Jiangxi: Jiangxi Normal University, 2021: 34-35. DOI: 10.27178/d.cnki.gjxsu.2021.000455.)

This work is partially supported by the Liaoning Natural Science Foundation (2022-MS-119), China University Industry-University-Research Innovation Fund (2022MU017), the National Natural Science Foundation of China (72271048).

HAN Yike, born in 1999, M. S. candidate. Her research interests include fault diagnosis, smart education.

XU Bin, born in 1980, Ph. D., associate professor, is a member of China Computer Federation. His research interests include artificial Intelligence, smart education, industrial intelligence, fault diagnosis.

ZHANG Shuo, born in 1988, Ph. D. candidate. His research interests include anomaly detection, machine learning, demand forecast.