

# 基于认知诊断的个性化试题推荐方法

朱天宇<sup>1)</sup> 黄振亚<sup>1)</sup> 陈恩红<sup>1)</sup> 刘淇<sup>1)</sup> 吴润泽<sup>1)</sup> 吴乐<sup>2)</sup> 苏喻<sup>3)</sup> 陈志刚<sup>1)</sup> 胡国平<sup>4)</sup>

<sup>1)</sup>(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

<sup>2)</sup>(合肥工业大学 计算机与信息学院, 合肥 230009)

<sup>3)</sup>(安徽大学 计算机科学与技术学院, 合肥 230039)

<sup>4)</sup>(科大讯飞股份有限公司, 合肥 230088)

**摘 要** 面向学生的个性化试题推荐是智能教育领域重要的研究课题, 现有的试题推荐工作大多采用协同过滤的方法或基于认知诊断的方法。然而, 协同过滤的试题推荐方法往往忽略了学生的学习状态(知识点掌握情况); 基于认知诊断的方法只能建模单个学生的学习状态, 不能利用相似学生的共性特征。针对以上问题, 本文提出一种基于学生知识点掌握程度的协同过滤试题推荐方法。该推荐方法分为三步: 第一步结合认知诊断模型, 根据学生已有的答题情况和试题知识点关联对学生的试题掌握水平进行建模, 第二步将学生的试题掌握水平用于概率矩阵分解预测学生答题情况, 第三步根据得分预测和试题难度向学生进行相应的试题推荐。该推荐方法同时考虑了被推荐学生学习的个性和考虑了群组学生学习的共性, 在保证试题推荐解释性的同时提高了试题推荐的可靠性。最后, 本文通过大量对比实验证明了该方法在进行学生试题推荐时能够保持精确性和可解释性。

**关键词** 协同过滤; 个性化试题推荐; 概率矩阵分解; 认知诊断; 知识点掌握

**中图法分类号** TP301

## 论文引用格式

朱天宇, 黄振亚, 陈恩红, 刘淇, 吴润泽, 吴乐, 苏喻, 陈志刚, 胡国平, 基于认知诊断的个性化试题推荐方法, 2016, Vol.39: 在线出版号 No.93

ZHU Tian-Yu, Huang Zhen-Ya, Chen En-Hong, Liu Qi, Wu Run-Ze, Wu Le, Su Yu, Chen Zhi-Gang, Hu Guo-Ping, Cognitive Diagnosis Based Personalized Question Recommendation, Vol.39: Online Publishing No.93

## Cognitive Diagnosis Based Personalized Question Recommendation

ZHU Tian-Yu<sup>1)</sup> Huang Zhen-Ya<sup>1)</sup> Chen En-Hong<sup>1)</sup> Liu Qi<sup>1)</sup> Wu Run-Ze<sup>1)</sup> Wu Le<sup>2)</sup> Su Yu<sup>3)</sup>  
Chen Zhi-Gang<sup>1)</sup> Hu Guo-Ping<sup>4)</sup>

<sup>1)</sup>(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027)

<sup>2)</sup>(School of Computer and Information, Hefei University of Technology, Hefei 230009)

<sup>3)</sup>(School of Computer Science and Technology, Anhui University, Hefei 230039)

<sup>4)</sup>(USTC iFLYTEK Co., Ltd., Hefei 230088)

**Abstract** Personalized question recommendation for students is a significant research direction in the domain of

本课题得到国家杰出青年科学基金(No. 61325010)、国家自然科学基金(No. 61403358)、中科院青年创新促进会会员专项基金(会员编号 2014299)、安徽省自然科学基金(No. 1408085QF110)资助。朱天宇, 女, 1993 年生, 硕士研究生, 主要研究领域为数据挖掘、推荐系统, E-mail: tianyu930331@gmail.com。黄振亚, 男, 1992 年生, 硕士研究生, 主要研究领域为教育数据挖掘、推荐系统, E-mail: huangzhy@mail.ustc.edu.cn。陈恩红(通讯作者), 男, 1968 年生, 博士, 教授, 计算机学会(CCF)会员(09201D), 主要研究领域为机器学习、数据挖掘、社会网络、个性化推荐系统, E-mail: cheneh@ustc.edu.cn。刘淇, 男, 1986 年生, 博士, 副教授, 计算机学会(CCF)会员(E200037367M), 主要研究领域为数据挖掘与知识发现、机器学习方法及其应用, E-mail: qiliuqi@ustc.edu.cn。吴润泽, 男, 1990 年生, 博士研究生, 主要研究领域为教育数据挖掘、认知诊断, E-mail: wrz179@mail.ustc.edu.cn。吴乐, 女, 1988 年生, 博士, 讲师, 计算机学会(CCF)会员(40748G), 主要研究领域为教育数据挖掘及知识发现、推荐系统、社交网络, E-mail: lewu.ustc@gmail.com。苏喻, 男, 1984 年生, 博士研究生, 主要研究领域为数据挖掘、图像识别, E-mail: firesysysy@163.com。陈志刚, 男, 1977 年生, 博士研究生, 中级职称, 计算机学会(CCF)会员(22262M), 主要研究领域为自然语言理解、人机交互、数据挖掘, E-mail: zgchen@iflytek.com。胡国平, 男, 1977 年生, 工学博士, 工程师, 计算机学会(CCF)会员(E200022261M), 主要研究领域为智能语音、语言核心技术研究, E-mail: gphu@iflytek.com。

intelligent education. Current studies depend on either collaborative filtering based methods or use the cognitive diagnosis models. Unfortunately, collaborative filterings ignore the knowledge states (e.g. skill proficiency) of students and cognitive diagnosis models can only model the knowledge state for a single student while the common features of similar students are neglected. To solve the problems above, we propose a three-step personalized question recommendation method(PMF-CD) which combines the complementary advantages of Probabilistic Matrix Factorization and Cognitive Diagnosis. Firstly, we model students' proficiency on each question by cognitive diagnosis models based on students' scores and question-skill correlation matrix. Secondly, we predict students' performance by probabilistic matrix factorization combined with students' skill proficiency. Finally, we recommend questions to students according to their predicted performance and the difficulty of questions. PMF-CD takes both individual and common study status of students into account, which ensures the interpretability and effectiveness of recommendation. Extensive experiments on real-world datasets demonstrate that PMF-CD performs well, and the recommendation is also interpretative.

**Key words** Collaborative filtering; Personalized Question Recommendation; Probabilistic Matrix Factorization; Cognitive Diagnosis; Skill Proficiency

## 1 引言

伴随着大数据时代的到来, 智能教育和在线教育的发展越来越快, 它们能够便捷的为学生提供试题练习, 帮助学生巩固学习的知识[1,2]。然而, 由于试题资源数量庞大, 学生想在有限的时间内练习所有的试题几乎是不可能的。因此, 如何帮助学生在浩如烟海的试题资源中找到合适的试题是一个重要的问题[3,4]。

近年来, 学者们尝试将推荐系统的相关技术应用到试题推荐的问题中[1,5]。类比于传统推荐系统在电子商务领域的应用, 他们将教育学对学生对象看作电子商务中的用户, 试题看作商品, 学生在试题的得分看作用户对商品的评分。因此, 学生试题的得分预测就可以使用推荐系统中常用的方法解决。这种方法简单易懂, 取得了良好的效果[5]。在传统的推荐系统中, 基于协同过滤的方法是应用最广的一类算法[6,7,8], 它又可分为两类: 一类是基于近邻的协同过滤方法, 另一类是基于模型的协同过滤方法。应用于试题推荐问题时, 基于近邻的协同过滤方法首先根据学生在试题上的答题记录计算学生之间的相似度找到最相似学生, 通过相似学生的得分对目标学生进行得分预测, 进一步根据预测的得分进行试题推荐[7]; 基于模型的协同过滤方法以矩阵分解[9]方法应用最为广泛, 它将学生和试题分解成一组潜在因子的影响因素, 构造学生和试题的低维矩阵, 来刻画学生和试题在低维空间中的表现程度, 并据此来预测学生在试题上的得分, 进一步

根据预测的得分进行试题推荐。[7]的实验结果和[10]的实验结果说明, 协同过滤的方法可以提高学生得分预测效果, 提高试题推荐的准确度。

教育数据挖掘中的认知诊断方法也可以用于学生个性化试题推荐。认知诊断方法首先通过各种不同的模型对学生进行能力建模, 建模所得到

表 1 试题知识点关联 Q 矩阵

试题	知识点				
	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
$j_1$	1	0	0	0	0
$j_2$	1	1	0	0	0

的参数均具有很强的解释性意义, 并在用于学生得分预测和试题推荐时表现较好。不同于传统推荐系统中的将学生视为用户和将试题视为商品对象, 在教育数据挖掘中, 有学者认为试题受到一组显式的知识点关联, 即试题会考查一组给定的、相关联的知识点, 并用 Q 矩阵表示[11]。且认为学生在试题上的表现受到学生对于试题所考察知识点的掌握程度的影响。表 1 为一个简单的 Q 矩阵的示例, 矩阵的每一行代表一个试题, 每一列代表一个知识点 (如一元二次方程), 从表 1 中可以看出试题  $j_2$  关联知识点  $s_1$  和  $s_2$ 。学生在试题上的得分受学生在试题所关联知识点上的掌握程度的影响, 对学生的能力 (知识点掌握程度) 的建模过程在认知心理学中被称为认知诊断模型[12]。认知诊断模型需要在建模学生的同时保证较好的解释性[11,13,14], 常见的认知诊断模型有项目反映理论 (Item Response Theory, IRT) 和 DINA 模型 (Deterministic Inputs, Noisy "And" gate model) [11,12,15]。其中, IRT[15]将学生描

述成一维的能力值，并结合试题参数(难度，区分度等)对学生进行建模。DINA 模型[11]结合 Q 矩阵，将学生描述成一个在多维知识点上的掌握向量，从而对学生建模。对于得到的学生模型，可以进一步预测试题得分从而进行试题推荐。虽然认知诊断模型通过结合教育学的领域知识(试题参数，Q 矩阵等)保证了其结果的解释性，然而由于学生的知识点掌握情况有一定的隐蔽性，无法直接对它进行测量，仅能通过学生的答题情况来推测其知识点掌握情况，推测过程中误差可能会导致学生知识点掌握程度的判断欠准确[16]。因此，仅通过学生知识点掌握情况进行学生试题推荐(得分预测)的误差相对较大。

总结起来，传统的协同过滤推荐方法和基于认知诊断的方法在应用于学生个性化试题推荐时各有优势而又各有不足。传统的协同过滤推荐方法在应用于个性化试题推荐场景时的不足在于协同过滤方法忽略了学生试题相关联的知识点信息，使得其在个性化试题推荐的场景下[9]：1. 难以保证推荐结果的可解释性，即无法向学生说明推荐某一试题的原因，学生不了解自己为什么被推荐某一试题(是因为对哪一知识点掌握的不足?)；2. 不能保证试题推荐结果的合理性，即所推荐的试题并不一定对学生的练习提高有所帮助。传统推荐算法仅根据学生的做题偏好给出推荐，这种推荐极易给出简单试题或难题，然而在个性化试题推荐的背景下，并不是向目标学生推荐越难的试题就越好，也不是向其推荐越简单的试题就越好[4,17]，应当推荐对学生的知识掌握有提升帮助的相关试题。例如在学习的提高阶段，学生可能需要有一定难度的试题进行练习以得到知识点掌握程度的提升，而在学习的巩固阶段，则需要中等难度的试题进行学习的巩固，然而基于协同过滤的推荐系统在推荐的过程中并不考虑学生之间知识点掌握程度的差异，推荐结果往往不够合理。另一方面，传统的基于认知诊断的试题推荐方法，虽然在充分利用学生的个性化知识点掌握程度的基础上保证了推荐结果的可解释性，却不能像协同过滤方法一样对相似学生的共性进行建模。在学生知识点掌握程度建模的过程中，各种建模方法都存在由于忽略了学生学习之间的共性，从而使得对数据隐含的信息挖掘不够充分的情况，并且学生知识点掌握程度作为学生的潜在隐藏属性难以准确测量，对学生的知识

点掌握程度建模可能欠准确，容易造成较大的建模和推荐误差。由于以上原因，基于认知诊断的试题推荐方法难以实际运用。

可以看出，目前个性化学生试题推荐研究中存在的主要问题有以下两个方面：1. 如何从学生、试题、知识点三者关系中精准对学生的个性化知识点掌握程度进行建模，使其能够合理应用于试题推荐？2. 如何有效结合学生个性化知识点掌握程度和相似学生的共性特征来解决个性化学生试题推荐问题，并同时保证推荐结果的合理性和可解释性？

为了解决以上问题，本文提出了一种基于认知诊断的个性化试题推荐方法 PMF-CD，希望通过融合两者的优势来弥补各自在试题推荐应用上表现出来的不足。PMF-CD 合理的结合了协同过滤和认知诊断各自的优势，能够生成精准且具有可解释性的学生个性化试题推荐。本文所提出的个性化试题推荐方法的输入为学生答题得分矩阵  $R$  和由教育专家标记的试题知识点关联矩阵(Q 矩阵)；该方法的第一步通过认知诊断模型对学生知识点掌握进行建模得到学生的个性化知识点掌握程度，并根据学生的个性化知识点掌握程度推测学生在试题上的掌握程度；第二步将学生在试题上的掌握程度作为先验参数应用于概率矩阵分解中(Probabilistic Matrix Factorization, PMF)[10,18]，推荐方法的第二步同时结合学生之间的学习共性和学生个人的试题掌握程度特性进行得分预测，通过在 PMF 分解中加入认知诊断因素(学生个性化试题掌握程度)，可以使得 PMF 分解的低维潜在因子包含学生的个性特征，提高结果的解释性；第三步通过调整推荐试题对于学生的难度，从而给学生进行个性化的试题推荐，保证推荐结果的合理性。因此，不同于传统的推荐系统仅围绕用户的兴趣点进行单一的推荐，PMF-CD 可根据学生的不同学习状态推荐不同难度的试题。最后，本文在两组真实的高中学生数学考试数据上的大量实验，验证了本文所提出的基于认知诊断的个性化试题推荐方法 PMF-CD 所生成推荐结果的有效性和可解释性。总之，本文所提出的个性化试题推荐方法的主要创新如下：

(1)据作者所知，PMF-CD 是首个可以结合 PMF 和认知诊断进行个性化试题推荐的方法。它从学生知识点掌握角度对学生建模，结合 PMF 对学生进行试题推荐，同时具备传统推荐方法所具备



的精确性, 以及认知诊断的可解释性。

(2) 该方法结合学生自身和其相似学生的做题情况进行试题推荐, 能考虑每个学生的学习情况个性和学生之间的共性, 提高了推荐结果的可靠性。

(3) PMF-CD 引入试题推荐难度范围参数, 使得该推荐方法可以向学生推荐不同难度范围的试题, 具有很好的扩展性和合理性。

本文剩余部分组织如下: 章节 2 将介绍试题推荐的相关技术。章节 3 将介绍问题定义和准备知识。章节 4 将介绍 PMF-CD 的框架和技术方法。章节 5 通过大量实验验证 PMF-CD 的有效性和可解释性。最后, 章节 6 将总结本文并提出展望。

## 2 相关工作

本章节将从推荐系统认知诊断两个部分介绍试题推荐方法的相关工作。

### 2.1 推荐系统

推荐系统是数据挖掘中重要的应用[19], 可以有效地在海量的资源中快速定位用户可能感兴趣的内容。在推荐系统中, 基于排序和基于评分预测的推荐方法是目前最流行的两种推荐方式。近年来, 有学者将基于排序和基于评分预测的推荐方法分别应用到教育的相关研究中, 在试题推荐等方面取得了一定的成果[5,18,20]。

对于这两种不同的推荐方式, 在众多的推荐算法中, 协同过滤 (Collaborative Filtering) 是其中应用最广泛的推荐算法之一[8], 也是目前研究最深入的推荐算法之一, 通过分析和目标用户的相似用户 (如用户评分相似、兴趣偏好相似、用户行为相似等) 来对目标用户进行推荐。

在基于排序的推荐方法应用于学生个性化试题推荐方面, Avi Segal 提出了联合协同过滤和社会选择理论的学生个性化建模方法 EDUrank[20]。该方法通过协同过滤找到待推荐学生的相似学生, 并根据相似学生在试题上的排序获得目标学生在不同试题上的难度排序。该方法可以用于帮助教师自动筛选适合学生的习题。

推荐系统中除了可以联系试题难度排序进行个性化试题推荐外, 还可以通过预测学生在试题上的得分 (predict student performance, PSP) 来进行个性化试题推荐[3]。应用于个性化试题推荐的

协同过滤算法又可以分为基于近邻的协同过滤和基于模型的协同过滤[21]。其中, 基于近邻的协同过滤方法首先根据学生在试题上的答题记录计算学生之间的相似度, 找到目标学生的相似学生群组, 通过相似学生的得分对目标学生进行得分预测, 进一步根据预测的得分进行试题推荐[7]。基于近邻的协同过滤方法自提出以来在许多数据挖掘领域的得到了较理想的效果, 且已经有学者将其用于教育领域的个性化推荐[22], 包括试题推荐、学习材料推荐、退学预警等。[23]提出一种在线学习背景下基于模糊树匹配用户学习材料个性化推荐; 除了向单个学生进行学习材料的个性化推荐之外, [24]提出了一种融合不同学生的学习兴趣并向一组学生进行个性化学习材料推荐的方法。但由于使用近邻学生表现预测学生表现时, 会丢失学生自身的一些学习特性, 基于近邻的协同过滤算法在教育应用中存在一定的局限性。基于模型的协同过滤算法在教育应用最为广泛的是矩阵分解[10], 它通过构造学生和试题的低维矩阵, 刻画学生和试题在低维空间中的表现程度, 并据此来预测学生在试题上的得分, 从而进行试题推荐。[9,10]表明, 矩阵分解已经可以取得较好的推荐精确度。近年来不断有学者提出概率矩阵分解在各种不同的场景下的改进方法, [25]提出在矩阵稀疏的情况下引入用户对项目的标签判断用户之间的相似性, 并通过为用户和项目分配特有的高斯分布来保证相似的用户可以分解得到相似的隐变量从而提升建模的效果; [26]提出了通过捕获丢失数据和观察过程之间的复杂依赖关系来对非随机缺失数据进行概率矩阵分解方法。但由于概率矩阵分解得到的潜在向量的含义难以解释, 使其在用于试题推荐时具有一定的局限性[9], 推荐结果的可解释性不强。

### 2.2 认知诊断

在认知心理学中, 认知诊断模型可以较好地知识点对学生的认知状态进行建模, 并在学生矫正计划、学习退学预警等应用中取得了良好的效果[13]。现存有大量的认知诊断模型, 可以从两个方面对这些认知诊断模型进行分类[12], 一方面, 认知诊断模型可分为离散型和连续型[27,12]; 另一方面, 认知诊断方法可以分为一维技能或多维技能的认知诊断[11,15]。在众多的认知诊断模型中, 应用最广泛的是一维连续认知建模的项目反应理论 (IRT)[15]以及多维离散认

知建模的 DINA 模型[11]。

其中 IRT 仅通过学生在试题上的答题情况将学生建模为一维连续能力值，并以这一个单一的能力值表示学生的综合能力。在 IRT 模型下，学生被刻画成一个具有单一能力值的对象，除题目自身的因素（区分度、难度、猜测度）之外学生在题目上的表现情况仅受到这单一的综合能力影响。然而在实际运用中，由于不同的试题考察范围的不同导致单一的学生综合能力难以刻画学生不同维度（知识点）上的能力差距，对此有学者提出了补偿性多维 IRT 模型 MIRT-C，和非补偿性多维 IRT 模型 MIRT-NC[12]。这两类模型均可以将学生的能力在多个维度上分别建模，是 IRT 模型建模所得到的学生能力值在维度上的改进。

在引入了知识点相关的能力维度后，有学者提出一个试题所考察的多个知识点之间的相互作用可以分为两类[12]，分别为“连接性的”知识点相互作用和“补偿性的”知识点相互作用。DINA 模型[11]属于多维离散认知建模，且该模型认为知识点间的相互作用是“连接性的”。DINA 模型除了使用学生在试题上的答题情况之外还引入了试题知识点关联矩阵（ $Q$  矩阵），通过联合更多的教育领域内特有的信息对学生的知识点掌握程度进行诊断并将学生建模成一个在多维知识点上的掌握向量，其中向量的每一维表示学生是否掌握了对应维度上的知识点。DINA 模型可以进一步结合试题上的猜测和失误参数，并预测学生在试题上的表现情况。继 DINA 模型之后多名学者提出了改进的 DINA 模型，包括 HO-DINA、P-DINA、G-DINA 等[27]。然而上述工作都只能给出学生知识点的离散化掌握程度（即知识点掌握或未掌握），在实际中，对知识点的掌握程度是 0.9 和 0.7 的两个学生在做题的表现上仍具有一定的差距，在 DINA 模型和其改进的模型中可能对这两个学生的建模结果均为掌握了该知识点，无法体现学生在知识点掌握程度上的细微差别。因此，DINA 模型在学生建模的精确度上具有一定局限性，无法量化诊断结果。

此外，IRT 模型和 DINA 模型只能处理离散数据，因此它对学生在客观题（得分非对即错）上的诊断具有良好的效果，但却不能精准地建模学生在主观题（得分为连续的值）上的表现。针对这一问题，有学者结合模糊系统，提出了一种针

对主客观题联合的学生建模方法[14]，并取得了良好的建模效果。

### 3 问题定义及准备知识

针对所要进行的具体工作，本章介绍了 PMF-CD 试题推荐方法的问题定义及准备知识。

在教育领域的试题推荐问题中，一方面，给定学生集合  $P = \{P_1, P_2, \dots, P_u\}$ ，试题集合  $J = \{J_1, J_2, \dots, J_v\}$ ，学生试题得分矩阵  $R = [r_{uv}]_{U \times V}$ ，其中， $r_{uv} = 1$  表示学生  $u$  答对试题  $v$ ， $r_{uv} = 0$  表示学生  $u$  答错试题  $v$ ；另一方面，对试题集合  $J$  所考察的所有知识点  $S = \{S_1, S_2, \dots, S_k\}$ ，试题与知识点关联情况由  $Q$  矩阵[11]定义： $Q = [q_{vk}]_{V \times K}$ ，其中， $q_{vk} = 1$  表示试题  $v$  考查知识点  $k$ ， $q_{vk} = 0$  表示试题  $v$  未考查知识点  $k$ ，试题对知识点考查情况由教育领域内的专家标注得到。则个性化试题推荐问题可定义如下：

**定义：**给定学生集合  $P$ 、试题集合  $J$ 、知识点集合  $S$ ，学生得分矩阵  $R$  和试题知识点关联  $Q$  矩阵，本文主要的研究目标为：1. 如何对给定学生  $P_u$  的做题记录和  $Q$  矩阵，同时考虑其个性学习状态和相似学生的共性信息，得到精确的得分预测？2. 如何进一步针对试题对于学生的难度信息，给出精确且具有可解释性的试题推荐？

表 2 给出了问题所需要的符号及对应描述。

表 2 试题推荐问题所涉及的符号及描述

符号	描述
$P$	学生集合
$P_u$	学生 $u$
$J$	试题集合
$J_v$	试题 $v$
$R$	学生试题得分矩阵
$r_{uv}$	学生 $u$ 在试题 $v$ 上的得分
$S$	知识点集合
$S_k$	知识点 $k$
$Q$	知识点考查矩阵
$q_{vk}$	试题 $v$ 对知识点 $k$ 的考查情况
$\alpha_u$	学生 $u$ 的知识点掌握向量
$\alpha_{uk}$	学生 $u$ 对知识点 $k$ 的掌握情况
$\eta_{uv}$	学生 $u$ 对试题 $v$ 的潜在作答情况
$M_u$	学生 $u$ 潜在特征向量

### $N_v$ 试题 $v$ 潜在特征向量

#### 3.1 学生知识点掌握情况诊断

为了得到学生的学习状态(学生个性状态), PMF-CD 使用 DINA 模型[11]诊断学生的知识点掌握程度。传统 DINA 模型通过联合学生的做题情况  $R$  矩阵和试题知识点关联  $Q$  矩阵对学生进行建模。每个学生  $P_u$  描述为一个知识点掌握程度向量  $\alpha_u = \{\alpha_{u1}, \alpha_{u2}, \dots, \alpha_{uK}\}$ , 其中每一维对应一个知识点,  $\alpha_{uk} = 1$  表示学生  $u$  掌握知识点  $k$ ,  $\alpha_{uk} = 0$  表示学生  $u$  未掌握知识点  $k$ 。

在已知学生  $P_u$  的知识点掌握向量  $\alpha_u$  的情况下, 对于学生  $P_u$  未作答的试题  $J_v$ , 可以根据下式(1)获得学生  $P_u$  对试题  $J_v$  的潜在作答情况:  $\eta_{uv}$

$$\eta_{uv} = \prod_{k=1}^K \alpha_{uk}^{q_{vk}} \quad (1)$$

其中,  $\eta_{uv} = 0$  表示学生  $P_u$  无法正确回答试题  $J_v$ , 若  $\eta_{uv} = 1$ , 则认为学生  $P_u$  可以正确回答试题  $J_v$ 。

此外, DINA 模型还引入试题参数(失误率(slip)和猜测率(guessing))来建模学生在真实状态下的答题情况。具体地, 学生  $P_u$  对试题  $J_v$  的响应由下式(2)给出:

$$P_v(\alpha_u) = P(R_{uv} = 1 | \alpha_u) = g_v^{1-\eta_{uv}} (1 - s_v)^{\eta_{uv}} \quad (2)$$

由于 DINA 模型中认为知识点之间的相互关系是“连接性的”, 其中失误率  $s_v$  定义为对于掌握了试题所需全部技能的学生仍然无法答对试题  $v$  的概率; 猜测率  $g_v$  定义为未掌握试题所需全部技能的学生答对试题  $v$  的概率。

DINA 模型使用 EM 算法来最大化式(2)的边缘似然[11], 从而得到  $\hat{s}_1, \hat{s}_2, \dots, \hat{s}_v$  和  $\hat{g}_1, \hat{g}_2, \dots, \hat{g}_v$  的参数估计; 并且学生  $P_u$  的知识点掌握向量  $\alpha_u$  可以通过最大化该学生试题得分后验概率来确定(式(3)), 从而得到学生的二分知识点掌握向量。

$$\begin{aligned} \hat{\alpha}_u &= \arg \max_{\alpha} P(\alpha | R_u) \\ &= \arg \max_{\alpha} L(R_u | \alpha, \hat{s}_v, \hat{g}_v) P(\alpha) \\ &= \arg \max_{\alpha} L(R_u | \alpha, \hat{s}_v, \hat{g}_v) \\ &= \arg \max_{\alpha} \prod_{v=1}^V P(R_{uv} | \alpha, \hat{s}_v, \hat{g}_v) \end{aligned} \quad (3)$$

在获得了学生知识点掌握情况后, 便可以结合学生知识点掌握情况和待推荐试题的知识点考查情况, 作为本文所提出的 PMF-CD 试题推荐方法的先验进行学生答题情况预测, 并用于推荐。

#### 3.2 概率矩阵分解(PMF)

概率矩阵分解(PMF)作为一种传统的预测方法, 被广泛的用于推荐系统[9]。将 PMF 应用于学生试题推荐时, 可以通过预测学生在试题上的得分来进行学生个性化试题推荐。首先通过将学生和试题分别对应到传统推荐系统中的用户和商品(或电影之类的待推荐项目), 学生的答题情况  $R$  矩阵及对应用户对物品的喜爱程度。

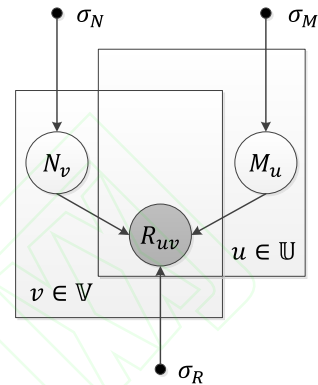


图1 学生成绩 PMF 图模型

图1 为将概率矩阵分解应用于学生得分矩阵分解时的图模型, PMF 将学生得分矩阵  $R$  分解成学生和试题的特征矩阵  $M$ 、 $N$ , 中  $M \in \mathbb{R}^{D \times U}$ 、 $N \in \mathbb{R}^{D \times V}$  且  $M_u$  和  $N_v$  分别表示特定学生  $u$  和试题  $v$  的潜在特征向量来刻画学生和试题低维空间下的表现。则学生试题得分  $R$  满足条件分布如下式(4)所示:

$$p(R | M, N, \sigma^2) = \prod_{u=1}^U \prod_{v=1}^V [N(R_{uv} | M_u^T N_v, \sigma^2)] \quad (4)$$

其中  $N(x | \mu, \sigma^2)$  是均值为  $\mu$ , 方差为  $\sigma^2$  的高斯分布概率密度函数。同时, PMF 假设学生和试题的特征向量满足均值为 0 的高斯分布:

$$\begin{aligned} p(M | \sigma_M^2) &= \prod_{u=1}^U N(M_u | 0, \sigma_M^2 I) \\ p(N | \sigma_N^2) &= \prod_{v=1}^V N(N_v | 0, \sigma_N^2 I) \end{aligned} \quad (5)$$

通过贝叶斯推论, 学生和试题的特征向量的后验概率推导如下:

$$\begin{aligned} p(M, N | R, \sigma_R^2, \sigma_M^2, \sigma_N^2) &\propto \\ & p(R | M, N, \sigma_R^2) p(M | \sigma_M^2) p(N | \sigma_N^2) \\ &= \prod_{u=1}^U \prod_{v=1}^V [N(R_{uv} | g(M_u^T N_v, \sigma_R^2))] \times \\ & \quad \prod_{u=1}^U N(M_u | 0, \sigma_M^2 I) \times \prod_{v=1}^V N(N_v | 0, \sigma_N^2 I) \end{aligned} \quad (6)$$

PMF 算法的图模型如图1所示, 通过公式(6), 可以仅通过已有的学生试题得分记录分解得到学生和试题的低维潜在因子向量, 能够很好地



刻画学生和试题的表现。

## 4 基于认知诊断的个性化试题推荐

### 方法 (PMF-CD)

针对基于协同过滤的学生个性化试题推荐方法和基于认知诊断的学生个性化试题方法各自所表现出来的不足,本文提出了一种新的学生个性化试题推荐方法 PMF-CD,该方法通过结合协同

过滤中的概率矩阵分解方法和认知诊断中的 DINA 模型,使得最终得到的个性化推荐试题同时具备认知诊断所得到的良好的可解释性和概率矩阵分解所得到的良好的推荐效果。

#### 4.1 PMF-CD方法整体框架

本章节介绍了本文所提出的 PMF-CD 的技术方案,PMF-CD 的整体流程图如图 2 所示。从图中可以看出该推荐方法主要分为 3 个步骤:

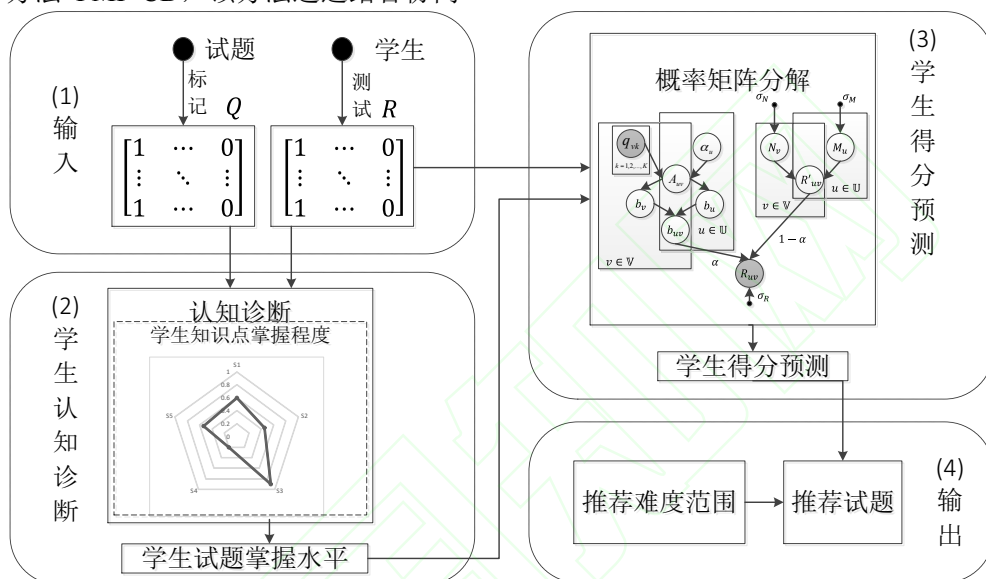


图 2 PMF-CD 整体框架

(1) 输入: 学生的做题情况  $R$  矩阵和试题知识点关联  $Q$  矩阵 (由领域内专家标注)。

(2) 学生认知诊断 (第一步): 使用 DINA 模型对学生知识点掌握程度进行建模,获得学生的知识点掌握程度,并根据学生在知识点上的掌握程度建模学生在试题上的掌握水平。

(3) 学生得分预测 (第二步): 结合 PMF 方法,并引入学生在试题上的掌握情况和待推荐试题知识点考察情况作为学生潜在因子和试题潜在因子的先验,据此预测学生试题得分。

(4) 输出 (第三步): 根据实际需求,设置待推荐试题相对于待推荐学生的题目难度范围  $[\beta_1, \beta_2]$ ,并筛选学生正确作答概率在推荐难度范围内的试题,从而向每个学生进行个性化试题推荐。

4.2.1 节将介绍 PMF-CD 试题推荐方法的第一步学生认知诊断;4.2.2 节将介绍第二步,结合学生认知诊断结果的学生得分预测;4.2.3 节将介绍该推荐方法的第三步,根据学生得分预测进行学生个性化试题推荐。

#### 4.2 PMF-CD方法具体做法

如前文所述,直接应用 DINA 模型和 PMF 进行得分预测均存在一定的不足,前者由于对内在因素衡量的不确定性导致预测准确性低,后者忽略了学生学习的个性且可解释性弱。为此,本文提出的 PMF-CD 方法结合 DINA 模型和 PMF 进行学生表现预测,作为试题推荐的基础。

##### 4.2.1 学生认知诊断 (第一步)

由于学生知识点掌握程度以及试题难度之间都有一定的差异,所以仅仅使用学生和学生作答信息通过概率矩阵分解获得学生和题目的低维特征并预测学生表现会损失一部分信息[9],损失的信息包括不同学生的知识点掌握情况的差异和不同试题的难度的差异等。为此,我们首先使用 DINA 模型对学生知识点掌握程度进行建模。

在使用 DINA 模型为学生知识点掌握程度进行建模时,针对 DINA 模型只能给出学生在知识点上的离散化的掌握程度 (只有掌握和未掌握两种) 的





解释性。

从图 3 可以看出, 学生的答疑预测由整体平均值、试题先验、学生先验和学生-试题交互四个部分组成, 每一个部分都解释了观测值的某一种属性, 则 PMF-CD 的优化目标可变为最小化函数公式:

$$E = \sum_{u=1}^U \sum_{v=1}^V I_{uv} (R_{uv} - \eta_{uv})^2 + \lambda_M \|M\|_{Fro}^2 + \lambda_N \|N\|_{Fro}^2 \quad (13)$$

其中,  $\lambda_M$ 、 $\lambda_N$  为模型的正则化系数。

总的来说, PMF-CD 方法在学生答题情况建模时, 通过结合学生的知识点掌握情况和试题难度等可解释信息, 使得该方法具备认知诊断模型的可解释, 在最大程度上从知识点层面排除学生的猜测和失误对答题情况建模的干扰、从试题层面调整试题难易程度对答题情况建模的干扰。

#### 4.2.3 根据学生得分预测进行试题推荐 (输出)

在已知学生潜在作答情况  $\eta_{uv}$  后, PMF-CD 个性化试题推荐方法的最后一步便是根据学生的潜在作答情况, 向不同的学生推荐个性化试题。本部分将描述 PMF-CD 根据学生潜在试题作答情况获得个性化试题推荐的具体方法。

不同于基于兴趣点的推荐系统, 在个性化试题推荐背景下, 并不是学生最感兴趣的试题、最难的试题或最简单的试题。为此, 在 PMF-CD 试题推荐方法中, 推荐试题的难度可以根据用户需求进行设定。试题难度指试题相对于学生的难度 [16], 以学生  $P_u$  正确回答试题  $J_v$  的概率作为试题  $J_v$  对于学生  $P_u$  的难度  $D_{uv}$ , 即:

$$D_{uv} = P(r_{uv} = 1 | P_u, J_v) \quad (14)$$

在进行试题推荐时, 设定推荐试题难度范围  $\beta_1$ 、 $\beta_2$  ( $\beta_1 < \beta_2$ ) 作为推荐试题难度边界, PMF-CD 可以根据学生的潜在作答情况  $\eta_{uv}$ , 向学生  $P_u$  从待推荐试题集合  $J_{rec}$  中向其推荐其正确作答概率在  $\beta_1$  到  $\beta_2$  之间的试题集合  $J_{rec\_u}$ :

$$J_{rec\_u} = \{J_v | J_v \in J_{rec}, D_{uv} \in [\beta_1, \beta_2]\} \quad (15)$$

其中  $\beta_1$  为学生正确作答概率下界、 $\beta_2$  为学生正确作答概率上界。因此, 在设定了相应的推荐试题难度范围后, PMF-CD 可以对每个学生推荐相应难度范围内的试题。例如, 将参数  $\beta_1$  设定为 0.3、参数  $\beta_2$  设定为 0.5 时, 本文提出的 PMF-CD 试题推荐方法可以从所有待推荐试题中找到对于该用户的

正确作答概率在 0.3 至 0.5 之间的试题进行推荐, 即公式(12)中  $\eta_{uv} \in [0.3, 0.5]$  的试题推荐给用户。

由于 PMF-CD 个性化试题推荐方法在进行试题推荐时, 有效地结合了学生自身学习情况的个性以及学生之间学习情况的共性, 则其推荐的试题能够体现学生当前的学习状态, 可以做到为学生推荐合适难度的个性化试题。

## 5 实验

### 5.1 数据集介绍

实验所用的两组数据均来自于某市高中学生数学考试的真实数据, 考试的数据真实的反映了学生对试题的掌握程度。其中, DATASET1 包含 259 个学生对 296 道试题的得分信息, 试题分别来自多次的单元性测试和一次综合性测试。这 296 道试题共考查如下 9 个知识点, 分别是: (S1)数系的扩充与复数、(S2)概率与随机变量、(S3)数列、(S4)三角函数与解三角形、(S5)立体几何、(S6)解析几何、(S7)常用逻辑用语、(S8)函数与导数、(S9)不等式; DATASET2 由 8819 个学生对 48 道试题的得分组成, 试题均来自于 3 次综合性测试。这 48 道题共考查 13 个知识点, 除了含有 DATASET1 中的 9 个知识点外, DATASET2 还包含如下 4 个知识点: (S10)集合、(S11)计数原理、(S12)平面向量、(S13)算法初步。关于两组数据集的相关统计信息如表 2 所示。图 4 所示为两组数据集中试题与知识点关联的 Q 矩阵明细。例如, 从图 4 中的左图可以看出 DATASET1 中第 1 题考查知识点 S8; 从图 4 中的右图可以看出 DATASET2 中第 2 题考查知识点 S9、S10 等信息。

表 3 数据集的相关统计

数据集	学生	试题	知识点
DATASET1	259	296	9
DATASET2	8819	48	13

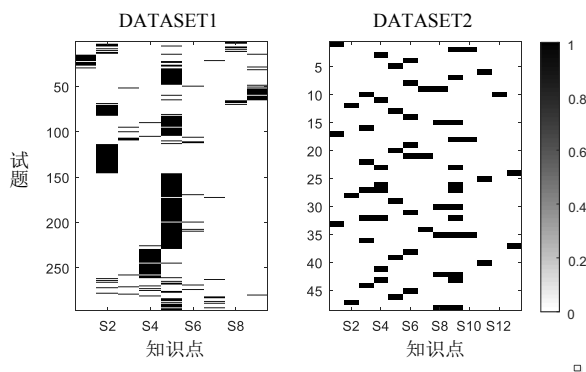


图4 Q矩阵

## 5.2 实验评价指标

### 5.2.1 精确率、召回率与 F1

为了衡量本文所提出的个性化试题推荐方法 PMF-CD 是否能更有效向学生推荐合适的试题, 本文使用精确率(*precision*)、召回率(*recall*)和 *F1* 指标来评价 PMF-CD 和其他对比方法在简单试题和困难试题推荐上的效果。具体地, 精确率、召回率和 *F1* 的定义如公式(16)所示:

$$\begin{aligned} precision &= \frac{TP}{TP + FP} \\ recall &= \frac{TP}{TP + FN} \\ F1 &= \frac{2 * precision * recall}{precision + recall} \end{aligned} \quad (16)$$

进行简单题推荐时, 公式(16)中的 *TP* 为被推荐试题中所有学生实际正确作答的数量; *FP* 为被推荐试题中所有学生实际未正确作答的数量; *FN* 为所有待推荐试题中学生实际正确作答的数量。

同样地, 在进行难题推荐时, *TP* 为被推荐试题中所有学生实际未正确作答的数量; *FP* 为被推荐试题中所有学生实际正确作答的数量; *FN* 为所有待推荐试题中学生实际未正确作答的数量。

### 5.2.2 正确作答率

由于精确率、召回率和 *F1* 值仅适用于评价推荐题目中最难的题目和最简单的题目, 若所进行的推荐超出学生的当前学习范围, 例如向初二的学生推荐高三的数学试题作为难题、推荐百位以内的加减法作为简单试题, 则很容易保证在极端情况下简单题和难题的推荐效果的精确率、召回率和 *F1* 值。在文章的第一部分介绍过, 不同于传统的推荐系统, 学生试题推荐这个推荐系统并不是向学生推荐越难的试题就越好, 也不是向学生推荐越简单的试题就越好。因此, 虽然精确率、召回率、*F1* 值可以从一个侧面体现出由 PMF-CD 算法推荐得到的个

性化推荐试题所得到的难题推荐相对于学生确实有一定难度, 所得到的简单题推荐学生容易答对, 但是这三个评价指标并不能保证 PMF-CD 推荐方法向学生推荐的中等难度试题是否真的对学生来说是否落在我们在推荐过程中所设定的难度范围。

为此, 在评价本文所提出的 PMF-CD 推荐的个性化难度的试题时, 本文提出了以正确作答率指标 *SR* 来评价被推荐试题对于学生的实际难度, 由设定难度范围后所得到的所有推荐试题的学生真实作答情况下的正确作答比例计算得出, 即:

$$SR = \frac{right}{total} \quad (17)$$

其中 *total* 为推荐试题总数, *right* 为推荐试题中学生正确作答的数量。*SR* 可以反映出所推荐的题目相对于学生的实际难度, 当推荐范围的上界和下届设置的都偏高时, 所推荐的题目对于学相对简单, 则学生的实际作答正确率应当偏高; 当推荐范围  $[\beta_1, \beta_2]$  的上界和下届设置都偏低时, 所推荐的题目相对于学生困难, 则学生的实际作答正确率应当偏低。

### 5.3 对比实验方法

为了验证 PMF-CD 试题推荐的效果, 本文将 PMF-CD 与如下实验方法进行对比:

**DINA 方法[11]:** 根据 DINA 诊断出的学生知识点掌握程度, 从待推荐试题集中选取学生知识点掌握较弱的知识点(困难试题推荐)和知识点掌握较强的知识点(简单试题推荐)的相关试题进行推荐。

**kNN 协同过滤方法[7]:** 通过学生已有试题得分, 计算学生间 Jaccard 相似度寻找答题情况最相似的学生, 再根据最相似学生的得分预测目标学生的得分, 进而根据预测得分的高低进行试题推荐。

**概率矩阵分解方法[10]:** 不引入学生个性学习状态和试题得分程度作先验(即公式(12)中的  $b_{uv}$ ), 仅通过传统 PMF 方法预测学生得分, 进而根据预测得分的高低进行试题推荐。特别地, PMF 方法也是 PMF-CD 方法在公式(12)中的特例(即  $\rho = 0$ )。

### 5.4 实验及结果分析

实验部分从以下五个方面进行实验并阐明实验结果:

(1)通过卡方检验, 从统计学的角度说明所提出的 PMF-CD 个性化试题推荐方法在学生潜在作答情况预测部分的预测有效性。

(2)对比试验, PMF-CD 相比于对比算法在难

题和简单题推荐效果是否有所提升？

(3)对于给定推荐试题难度范围  $[\beta_1, \beta_2]$ ，PMF-CD 推荐的试题是否符合推荐难度的要求？

(4)参数  $\rho$  如何影响 PMF-CD 的推荐效果？

(5)通过案例分析说明对于学生的实际推荐试题，是否真实反映了他当前的学习状态（符合他知识点掌握程度）？

#### 5.4.1 PMF-CD 得分预测卡方检验

由于无论是学生学习的个性和共性都难以直接测量和检验，难以直接确定建模所得到的学生知识点掌握向量和隐含知识点掌握向量的低维潜在因子是否准确，更难以直接根据所获得的知识点掌握向量和潜在因子向学生进行试题推荐。为了解决这个问题，在所提出的个性化试题推荐方法 PMF-CD 中，首先根据建模的结果模拟学生在测试集数据上的潜在作答情况  $\eta_{uv}$ ，并根据模拟得到的  $\eta_{uv}$  进行试题推荐。

因此在进行对比实现之前，增设 PMF-CD 推荐方法中得分预测部分的卡方检验，从统计的角度证明所以出的个性化试题推荐方法 PMF-CD 在成绩预测部分的有效性。

由于学生试题得分 R 矩阵数据分布为离散型，此处我们将 PMF-CD 方法中预测得分大于等于 0.5 的情况视为预测学生对该题的回答正确，小于 0.5 的情况视为预测学生对该题的回答错误。

#### 假设检验：

H0: PMF-CD 方法预测学生潜在作答情况  $\eta_{uv}$  和学生实际作答情况相互独立，PMF-CD 方法预测学生得分和实际学生得分无关。

H1: PMF-CD 方法预测学生潜在作答情况  $\eta_{uv}$  和学生实际作答情况不相互独立，PMF-CD 方法预测学生得分和实际学生得分相关。

置信度： $\alpha = 0.005$

我们在 DATASET1 和 DATASET2 中分别随机选取 20% 的学生在 40% 的试题上的作答情况进行卡方验证，统计结果如下表 4 所示。

通过计算两组数据集下抽样得到的  $\chi^2$  值，可得

$\chi^2_{DATASET1} = 1112.793$ 、 $\chi^2_{DATASET2} = 3380.656$ ，根据卡方检验 p 值表可以得到  $p_{DATASET1} < 0.005$ 、

$p_{DATASET2} < 0.005$ ，因此我们拒绝原假设 H0，接受假设 H1，即通过统计学上的卡方验证可以看出 PMF-CD 方法预测学生得分和实际学生得分相关，即验证了本文所提出的基于认知诊断的个性化试题推荐方法在学生得分预测部分的有效性。

表 4 PMF-CD 对两种作答情况的预测结果

(a) DATASET1			
分组	$\eta_{uv} = r_{uv}$	$\eta_{uv} \neq r_{uv}$	合计
$r_{uv} = 1$	3595(3078)	338(855)	3933
$r_{uv} = 0$	1207(1724)	996(478)	2203
合计	4802	1334	6136
(b) DATASET2			
分组	$\eta_{uv} = r_{uv}$	$\eta_{uv} \neq r_{uv}$	合计
$r_{uv} = 1$	20598(18654)	2529(4474)	23127
$r_{uv} = 0$	6435(8379)	3954(2001)	10389
合计	27033	6483	33516

\*括号内为理论频数

#### 5.4.2 简单题和难题推荐对比试验

在比较试题推荐效果的实验中，实验分别从简单试题推荐和困难试题推荐两个角度，将 PMF-CD 方法与 5.3 节所描述的 3 种对比实验方法进行对比试验。简单题推荐的做法为向目标学生  $P_u$  推荐待推荐试题集中对该学生最简单的试题（组）（即该学生潜在作答情况  $\eta_{uv}$  最高的试题（组）），难题推荐的做法为向学生  $P_u$  推荐待推荐试题集中对该学生最难的试题（组）（即该学生潜在作答情况  $\eta_{uv}$  最低的试题（组））。具体地，在两组数据集中，分别在学生和试题两个维度进行交叉验证。分别随机选取 20% 的学生作为测试集学生进行试题推荐，其余 80% 学生作为训练集。而对于试题，本文依次选择试题测试集(待推荐试题总量)的比例为 30%、40%、50%，其余的试题作为训练集。

以总试题的 30% 做测试集的实验为例。实验时以全部的学生在 70% 测试集中的试题作答情况获得学生知识点掌握程度，再以 80% 的测试集学生在全部试题上的作答情况获得试题的失误差和猜测率。这样的交叉验证划分保证了测试集学生在测试集试题上的作答情况不用于任何的参数训练。



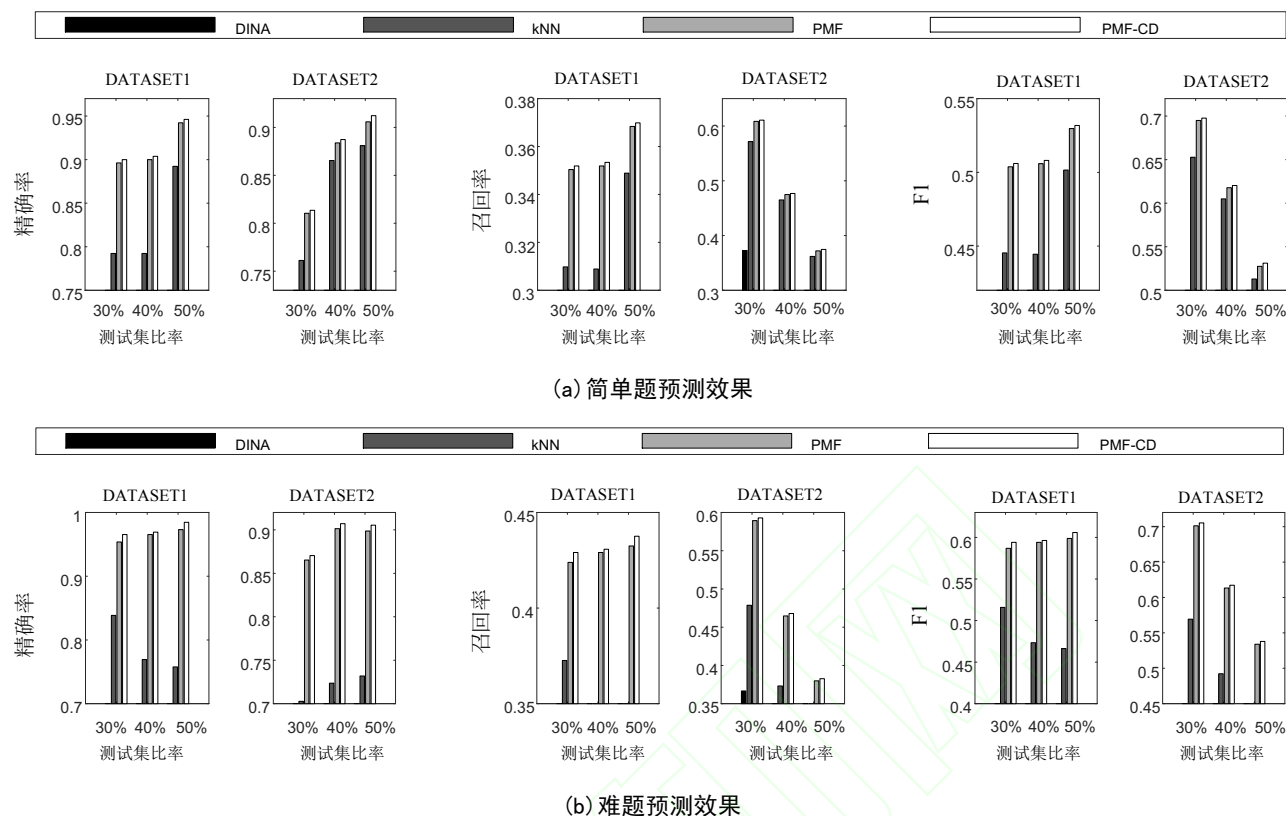


图5 不同推荐方法在不同数据集下推荐效果

表5 不同推荐方法在不同数据集下推荐效果的准确率、召回率和F1

	方法	简单题预测效果						难题预测效果					
		DATASET1 测试集试题比率			DATASET2 测试集试题比率			DATASET1 测试集试题比率			DATASET2 测试集试题比率		
		30%	40%	50%	30%	40%	50%	30%	40%	50%	30%	40%	50%
精确率	DINA	0.3231	0.3654	0.3577	0.4965	0.5445	0.5964	0.5192	0.5077	0.5000	0.5383	0.6066	0.6361
	kNN	0.7923	0.7923	0.8923	0.7612	0.8654	0.8809	0.8385	0.7692	0.7577	0.7027	0.7236	0.7319
	PMF	0.8962	0.9000	0.9423	0.8104	0.8838	0.9058	0.9538	0.9654	0.9731	0.8653	0.9014	0.8987
	PMF-CD	<b>0.9000</b>	<b>0.9038</b>	<b>0.9462</b>	<b>0.8135</b>	<b>0.8873</b>	<b>0.9122</b>	<b>0.9654</b>	<b>0.9692</b>	<b>0.9846</b>	<b>0.8704</b>	<b>0.9072</b>	<b>0.9054</b>
召回率	DINA	0.1263	0.1429	0.1398	0.3728	0.2926	0.2450	0.2308	0.2256	0.2222	0.3667	0.3128	0.2688
	kNN	0.3098	0.3089	0.3489	0.5715	0.4651	0.3619	0.3726	0.3419	0.3368	0.4787	0.3732	0.3094
	PMF	0.3504	0.3519	0.3684	0.6085	0.4749	0.3721	0.4239	0.4291	0.4325	0.5894	0.4648	0.3799
	PMF-CD	<b>0.3519</b>	<b>0.3534</b>	<b>0.3699</b>	<b>0.6108</b>	<b>0.4768</b>	<b>0.3747</b>	<b>0.4291</b>	<b>0.4308</b>	<b>0.4376</b>	<b>0.5929</b>	<b>0.4679</b>	<b>0.3827</b>
F1	DINA	0.1816	0.2054	0.2010	0.4259	0.3806	0.3473	0.3196	0.3124	0.3077	0.4362	0.4128	0.3779
	kNN	0.4454	0.4445	0.5016	0.6528	0.6050	0.5130	0.5159	0.4734	0.4663	0.5694	0.4923	0.4349
	PMF	0.5038	0.5059	0.5297	0.6950	0.6178	0.5275	0.5869	0.5941	0.5988	0.7012	0.6134	0.5340
	PMF-CD	<b>0.5060</b>	<b>0.5081</b>	<b>0.5318</b>	<b>0.6977</b>	<b>0.6203</b>	<b>0.5312</b>	<b>0.5941</b>	<b>0.5964</b>	<b>0.6059</b>	<b>0.7053</b>	<b>0.6173</b>	<b>0.5380</b>

依次划分实验训练集和测试集后, 本文使用上述4种实验方法在两组数据集中分别向每个测试集中学生推荐5道困难试题和5道简单试题, 并分别计算4种方法的精确率、召回率与F1值, 通过比较三个参数来评价四种不同的推荐方法的试题(简单题/难题)推荐效果。

实验中, 本文设定PMF-CD方法分解的 $M$ 、 $N$ 低维矩阵的维度为10, 为防止PMF-CD模型在训练时出现过拟合现象[10], 本文设定正则化系数 $\lambda_M = \lambda_N = 0.01$ , 模型中的参数 $\rho = 0.4$ (将在5.6参数 $\rho$ 的设置中通过试验详细阐述试验中设置 $\rho = 0.4$ 的原因)。试题推荐实验对比结果如图5和

表 5 所示（表 5 为图 5 具体数据），从中可以看出以下对比信息：

(1)在 DATASET1 和 DATASET2 下所有测试集比例的试验中，无论是对简单试题推荐，还是困难试题推荐，引入学生个性学习状态和试题得分程度作为先验的 PMF-CD 方法的试题推荐效果均优于其他算法。说明 PMF-CD 在简单题和难题的推荐上相比于其他的三种对比试验方法加有效。

(2)且相比于传统 PMF 算法，在测试集比例较小时，PMF-CD 方法的优势不是特别明显，但是随着训练集的减小，测试集的增大，PMF-CD 的优势体现出来，这说明在 PMF 分解中加入学生个性化学习状态  $b_w$ ，可以使得 PMF 分解出的低维潜在因子  $M, N$  在包含学生之间共性的学习状态的同时加入学生的个性特征，说明在综合了学生的学习个性和学生之间的学习共性之后，能更精确的对学生进行建模，提高学生个性化试题推荐效果。

(3)从图 5 和表 5 的实验结果中还可以看出，随着训练数据量的下降，基于 kNN 协同过滤的试题推荐方法效果呈明显的下降趋势，这可能是由于随着训练数据比例的下降，kNN 算法难以通过相似度

计算方法精准找出学生之间的相似程度最大的学生而影响了最终的试题推荐效果，但本文所提出的个性化试题推荐方法 PMF-CD 在包含相似学生的学习状态之外，还包含待推荐学生自身的学习个性，推荐效果较好。

(4)此外，DINA 模型的试题推荐的效果很不稳定，在训练集比例较高时的推荐效果较差，这可能是由于训练数据的减少使得对学生知识点掌握程度诊断的误差有明显增大而导致的。且由于 DINA 方法进行试题推荐的效果明显较差，本文在部分实验测试集的图上没有进一步展示该方法的结果，具体结果可参见表 5。可以看出，虽然学生的知识点掌握情况作为学生的隐藏属性难以衡量，诊断所得到的学生知识点掌握状态也在试题推荐时表现出的效果相比另外两种对比方法差，但当我们将从学生知识点掌握程度得到学生试题掌握程度作为先验信息，并加入 PMF 方法形成本文所提出的基于认知诊断的个性化推荐方法 PMF-CD 时，却提高了在简单题和难题上的推荐效果。

从对比试验的结果可以看出，PMF-CD 在简单题和难题推荐准确性上的优势。

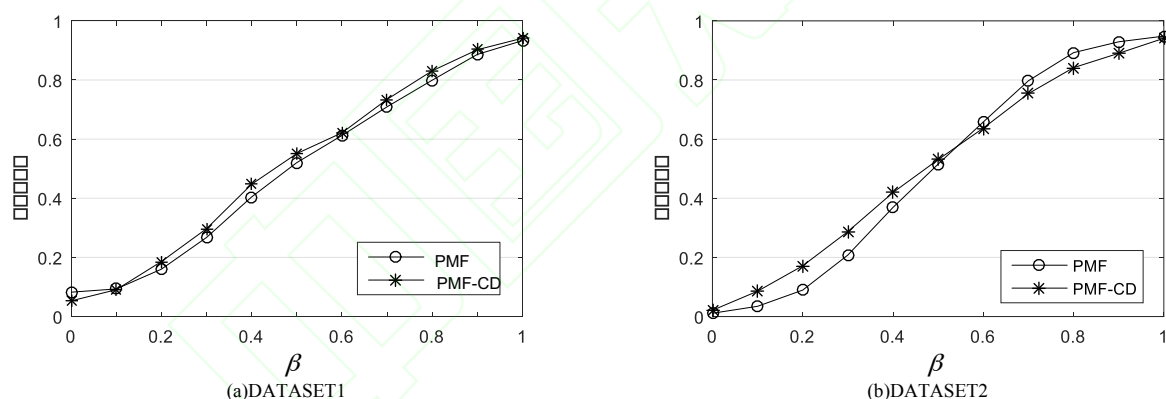


图 6 不同推荐难度  $\beta$  对测试集中推荐效果影响

#### 5.4.3 推荐试题难度参数 $\beta$ 的设置

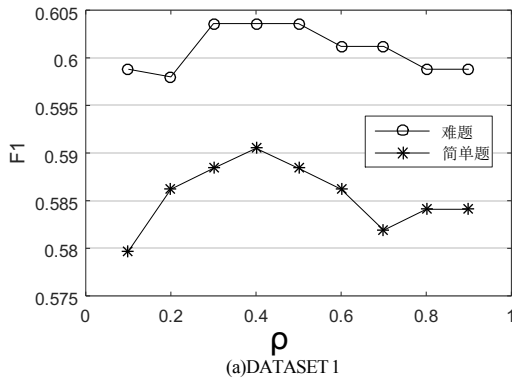
在本文所提出的个性化试题推荐方法的最后一步中，可以设定待推荐试题的难度范围。为了验证对于给定推荐试题难度范围  $[\beta_1, \beta_2]$ ，PMF-CD 推荐的试题是否符合推荐难度的要求，进行参数  $\beta$  的不同设置下学生相应 SR 变化趋势的实验。

本文使用测试集上的学生的正确作答率(SR)来验证 PMF-CD 方法是否能够有效地为学生推荐合适的试题，并选用不加学生个性化知识点掌握程度的 PMF 方法作为对比方法。具体地，本文将推荐试题的难度  $\beta$  依次划分为 11 个范围，从 0 到 1（包

括 0 和 1）每间隔 0.1 取一次  $\beta$  并以  $[\beta-0.1, \beta+0.1]$  作为推荐题目难度范围进行试验。实验中依次提高推荐试题的难度，在测试集中根据上述两种算法的预测得到的学生潜在作答情况筛选得到对每个学生在该难度范围下的个性化推荐试题集，计算测试集上学生对于推荐试题的正确作答率(SR)并观察 SR 随  $\beta$  的变化情况，实验结果如图 6 所示。

从图 6 可以看出，随着推荐难度的减小，学生对所推荐的试题正确作答率(SR)不断提高，说明 PMF-CD 能够根据推荐试题难度的需要，推荐适合学生的难度的试题。例如，设定  $\beta$  为 0.4 时，学生在 PMF-CD 算法推荐的试题上的实际正确作答率

接近 0.4。此外,从图 6(b)中可以看出,在样本量更大的 DATASET2 中,根据不同设定难度的要求,PMF-CD 方法的试题推荐效果稳步上升,能够更接近学生做题水平,而传统 PMF 方法的试题推荐效果存在一定的误差,例如,图 6(b)中设定试题难度  $\beta$  为 0.2 时,学生在 PMF 算法推荐的试题上的正确作答率(SR)约为 0.1。另外,PMF-CD 在图 6(a)的 DATASET1 上的推荐效果与 PMF 算法类似,这可能是 DATASET1 的数据量较小,而导致 PMF-CD 算法的优势不够明显。



通过参数  $\beta$  设置的实验可以看出,本文所提出的基于认知诊断的个性化试题推荐方法 PMF-CD 在引入了学生个性化知识点掌握情况后,相对于不引入学生个性化学习状态的情况可以更加精准的向每个学生推荐所设定推荐难度范围内的试题。该方法在结合了学生学习的个性和学生之间的学习共性之后,可以更加准确的进行学生的潜在作答情况预测,从而使得基于这种更准确的潜在作答情况预测得到的个性化试题推荐的结果更加可靠且可解释性好。

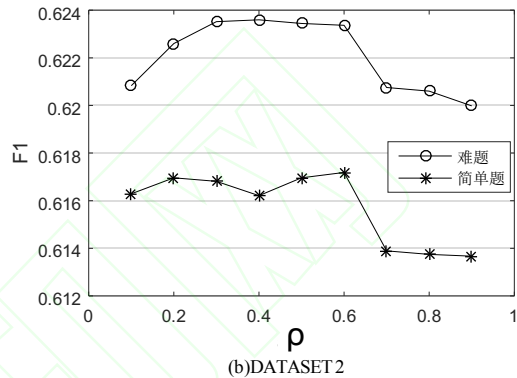


图 7  $\rho$  值对 PMF-CD 效果的影响

#### 5.4.4 参数 $\rho$ 的设置

本文所提出的 PMF-CD 模型中,参数  $\rho$  作为调整 PMF 方法和认知诊断程度的权重参数出现,且  $\rho \in [0,1]$ 。 $\rho$  越大,则表示学生得分越接近学生个性化学习状态; $\rho$  越小,则表示学生得分越接近学生之间的共性学习状态。特别地,当  $\rho=0$  时,该方法退化为 PMF。

在参数  $\rho$  的设置实验中,我们分别在 DATASET1 和 DATASET2 中选取 20% 的学生在 40% 的试题上的作答情况作为测试集进行不同的  $\rho$  对本文所提出的 PMF-CD 个性化试题推荐方法在简单题和难题推荐上的推荐效果进行实验。

图 7 显示了 PMF-CD 分别在两个数据集下进行简单试题和困难试题推荐时,算法的推荐结果效果(F1 值)与参数  $\rho$  值的变化关系。从图 7 可以看出, $\rho$  值的选取对 PMF-CD 方法的试题推荐效果有明显的影响。由图 7(b)可以看出,对于 DATASET2,  $\rho=0.6$  可以使得 PMF-CD 进行困难试题推荐的效果最好;  $\rho \in [0.4, 0.6]$  可以保证 PMF-CD 进行简单试题推荐有较好的效果。仔细分析图 7 不难发现,在  $\rho=0.6$  上升到  $\rho=0.7$  时在 DATASET1 中的难题推荐下有略微的下降,而在比 DATASET1 样本数量多的 DATASET2 中,此处

的下降非常明显,我们认为在  $\rho=0.6$  这一点周围该方法表现出来的稳定性不够;反之在  $\rho=0.4$  的周围,在 DATASET1 上或是在 DATASET2 上、在简单题预测表现和难题预测表现上其 F1 值都相对平稳且表现较好。综合考虑  $\rho$  取值不同时 PMF-CD 推荐效果的自身表现和其潜在的稳定性,这便是本文在 5.5 部分选定  $\rho=0.4$  作为实验的模型参数的原因。

然而,图 7(a)中显示,在 DATASET1 上,  $\rho$  值的选取对于 PMF-CD 在简单试题上的影响效果不明显,这可能是由于 DATASET1 数据量较小,从而影响了算法的效果。此外,我们还能发现,  $\rho$  值的选择对于 PMF-CD 进行简单试题推荐的影响相比于困难试题推荐较弱,这可能是由于简单试题对于学生学习状态的反映不够明显,而困难试题能够比较准确地反映学生的学习状态。

#### 5.4.5 案例分析

本节通过实际案例分析展示了 PMF-CD 对于两个学生的试题推荐情况。图 8 显示了在 DATASET1 上学生 A、B 对于 9 个知识点的掌握程度。可以看出,学生 A 在 S1、S4、S8 等知识点上掌握较好,知识点 S3、S9 上掌握较差;学生 B 在知识点 S1、S7 上掌握较好,在知识点 S2、S5 上掌握较差。



在测试数据集中, 学生 A 和学生 B 的正确作答率均为 50%, 当选取难度在 0.3-0.5 之间的试题进行推荐时, PMF-CD 向学生 A 推荐的试题为推荐集中 13 题和 201 题、向学生 B 推荐的试题为推荐集中 7 题、132 题和 201 题。其中, 132 题考查的是 S2 知识点, 由图 8 可以看出学生 B 的 S2 知识点较弱, 学生 A 的 S2 知识点在中等水平。由于选取的试题难度为 0.3-0.5 之间属于难度较高的试题, 132 题相对于学生 A 属于难度一般的试题, 相对于学生 B 属于难度较高的试题。

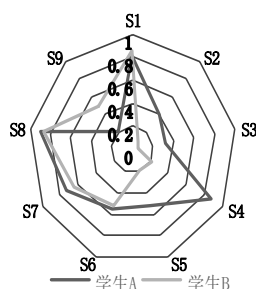


图 8 学生知识点掌握程度示例

从该案例分析可看出, 本文提出的个性化试题推荐方法 PMF-CD 在推荐难度较高的试题时, 可以针对每个学生的个性化学习状态向其推荐相应的试题, 试题推荐结果具有很强的可解释性。

### 5.5 实验结果讨论

从五个部分的实验结果可以看出, 本文所提出的基于认知诊断的个性化试题推荐方法 PMF-CD 无论是在简单题/难题的推荐效果上还是在所推荐试题难度和选择的推荐难度范围的吻合度上均有较好的效果, 这是因为该方法同时考虑了学生学习状态的个性(知识点掌握情况)和学生之间的学习共性(协同过滤方法)从而提高了推荐效果的可靠性和可解释性。

由于本文所使用的数据集为高中学生数学测试的真实数据, 且两组数据集中学生分别参加了所有的考试, 故两组数据集的填充率为 100%, 而在实际的场景中, 由于高中数学的试题总量极其庞大, 每个学生以其有限的精力只能完成题海中试题的少部分, 同时每道题也只会有一部分学生进行作答, 故可观测的学生数据往往是稀疏的。由本文的实验结果可以看出, PMF-CD 因为结合了学生学习的个性和共性, 在获得学生低维潜在因子时加入了其知识点掌握情况作为先验参数, 这样结果既有 PMF 分解的学生学习共性, 也包含了

由 DINA 模型得到的学生知识点掌握程度个性, 就使得 PMF-CD 方法既避免了仅使用知识点掌握程度进行推荐时因为训练集小而对学生知识点掌握程度测量不准的问题, 也避免了仅使用 PMF 方法时可解释性差的欠缺, 从 5.4.5 小节的案例分析可以看出本文提出的基于认知诊断的个性化试题推荐方法使得推荐结果的可靠性和可解释性增强。

## 6 结论和展望

针对传统认知诊断模型和传统 PMF 算法在学生试题推荐方面的缺陷, 即前者忽略了学生之间的学习个性、后者忽略了学生的学习个性, 本文提出了一种结合概率矩阵分解和认知诊断的个性化试题推荐方法 PMF-CD。该方法的输入为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); PMF-CD 的第一步结合认知诊断模型对学生知识点掌握程度进行建模, 并联合试题的知识点考察情况获得学生在试题上的掌握程度; 第二步将学生的个性试题掌握程度作为 PMF 算法的先验信息使用, 进行学生的得分预测, 在考虑学生群体之间学习的共性的同时加入了被推荐学生个人的学习状态个性; 第三步根据给定推荐试题的难度范围 $[\beta_1, \beta_2]$ , 筛选出适合目标学生做题难度的试题形成待推荐试题集, 并向学生进行推荐。文章的最后, 通过大量实验证明了本文所提出的基于认知诊断的个性化试题推荐方法 PMF-CD 的合理性和可解释性。

本文所提出的推荐方法, 将所有学生的个性化先验的权重参数  $\rho$  视为同等重要, 而在实际的学习过程中, 有些学生的学习个性强, 有些学生的学习个性弱。在进行 PMF-CD 进行试题推荐时, 若考虑学生学习的个性强度, 则会让推荐结果的可靠性更高, 这也是未来的研究方向之一。

### 参考文献

- [1] Vukicevic M, Jovanovic M, Delibašić B, Suknovic M. Recommender system for selection of the right study program for higher education students//Markus Hofmann, Ralf Klinkenberg. RapidMiner: Data Mining Use Cases and Business Analytics Applications. New York, USA: CRC Press. 2013: 145.
- [2] Ruth C. C, Richard E. M. E-learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning. Third Edition. San Francisco, USA: John Wiley & Sons Press, 2011.
- [3] Klačnja-Milićević, Aleksandra, Vesin, B, Ivanović, M. Budimac, Z. E-learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, 2011, 56(3):885-899.

- [4] Chen Xiao-Hong. Research and practice of optimizing design of mathematical homework in senior high school[Master Thesis]. Soochow University, Suzhou, 2010. (in Chinese)  
(陈晓红. 优化高中数学作业设计的实践与研究[硕士学位论文]. 苏州大学, 苏州, 2010.)
- [5] Walker, A., Recker, M. M., Lawless, K., Wiley, D. Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence in Education*. 2004, 14(1), 3-28.
- [6] Paul R, Neophytos I, Mitesh S, Peter B, John R. GroupLens: an open architecture for collaborative filtering of netnews//Proceedings of the 1994 ACM conference on Computer supported cooperative work. Association for computing machinery, New York, USA, 1994:175-186.
- [7] Greg L, Brent S, Jeremy Y. Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing*, Institute of electrical and electronics engineers, 2003, 7(1): 76-80.
- [8] Zheng Zi-Bin, Ma Hao, Michael R. L., Irwin K. Wsrec: a collaborative filtering based web service recommender system//IEEE International Conference on Web Services. IEEE, Los Angeles, USA, 2009: 437-444.
- [9] Koren Y, Robert B, Chris V. Matrix factorization techniques for recommender systems, *IEEE COMPUTER*. 2009, 42(8):30-37.
- [10] Ruslan S, Andriy M. Probabilistic matrix factorization.//Advances in neural information processing systems. Vancouver, Canada. 2007: 1257-1264.
- [11] De La Torre J. DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 2009, 34(1): 115-130.
- [12] DiBello L. V, Roussos L. A, Stout W. 31A Review of cognitively diagnostic assessment and a summary of psychometric models. //C.R Rao, S. Sinharay. Handbook of statistics 26: Psychometrics. Oxford, UK: Elsevier B.V., 2007: 979-1030.
- [13] Jacqueline P. L, and Mark J. G. Cognitive diagnostic assessment for education: Theory and applications. New York, USA: Cambridge University Press, 2007.
- [14] Wu Run-Ze, Liu Qi, Liu Yu-Ping, Chen En-Hong, Su Yu, Chen Zhi-Gang, Hu Guo-Ping. Cognitive Modelling for Predicting Examinee Performance. // The 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015:1017-1024.
- [15] Xitao F. Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological measurement*, 1998, 58(3): 357-381.
- [16] Dai Hai-Qi, Zhang Feng, Chen Xue-Feng. Psychological and Educational Measurement. Third Edition. Guangzhou, China: Jinan University Press, 2011(in Chinese)  
(戴海奇, 张峰, 陈雪枫. 心理与教育测量. 第三版. 广州, 中国: 暨南大学出版社. 2011)
- [17] Wang Hua-Chun. Reflections on the junior high school math homework design. *New Curriculum·High School*, 2013, 10:153(in Chinese)  
(王化春. 对初中数学作业设计的几点思考. 新课程·中学, 2013, 10:153)
- [18] Nguyen Thai-Nghe, Lucas Drumond, Tomas Horvath, Atrus Krohn-Grimberghe, Alexandros Nanopoulos, Lars Schmidt-Thieme. Factorization techniques for predicting student performance.//Olga C.Santos, Jesus G. Boticario, Educational Recommender Systems and Technologies: Practices and Challenges, Hershey, USA: IGI Global, 2011: 129-153.
- [19] Gediminas Adomavicius, Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6):734-749.
- [20] Shani, Guy, and Bracha Shapira. Edurank: A collaborative filtering approach to personalization in e-learning. // Proceedings of the 7th International Conference on Educational Data Mining. London, UK, 2014: 68-75.
- [21] Jang E E. A framework for cognitive diagnostic assessment.// Chapple Carol A., Chung Yoo-Ree, Xu Jing. Towards adaptive CALL: Natural language processing for diagnostic language assessment. Iowa, USA: Iowa State University Press, 2008: 117-131.
- [22] Andrew E. W, Mimi M. R, Kimberly L, David W. Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence and Education*, 2004, 14(1): 1.
- [23] Wu, Dianshuang, Jie Lu, and Guangquan Zhang. A fuzzy tree matching-based personalized e-learning recommender system. *IEEE Transactions on Fuzzy Systems*, 2015, 23(6): 2412-2426.
- [24] Dwivedi, Pragya, and Kamal K. Bharadwaj. E-learning recommender system for a group of learners based on the unified learner profile approach. *Expert Systems*, 2015 32(2): 264-276.
- [25] Wu Le, Chen En-Hong, Liu Qi, Xu Lin-Li, Bao Teng-Fei, Zhang Lei. Leveraging tagging for neighborhood-aware probabilistic matrix factorization. //Proceedings of the 21st ACM international conference on Information and knowledge management. Maui, HI, USA, 2012:1854-1858.
- [26] Jose M. H., Neil H. Zoubin G. Probabilistic matrix factorization with non-random missing data. //Proceedings of the 31st International Conference on Machine Learning. Beijing, China, 2014:1512-1520.
- [27] Zhang Xiao, Sha Ru-Xue, Research Advance in DINA Model of Cognitive Diagnosis, *China Examinations*, 013 (1): 32-37. (in Chinese)  
(张潇, 沙如雪. 认知诊断 DINA 模型研究进展. 中国考试, 2013 (1): 32-37.)
- [28] Toon C., Mykola P. Introduction to the special section on educational data mining. *ACM SIGKDD Explorations Newsletter*, 2012, 13(2): 3-6.
- [29] Mohsen J, Martin E. A matrix factorization technique with trust propagation for recommendation in social networks. //Recsys Proceedings of the Fourth Acm Conference on Recommender Systems, New York, USA, 2010:135-142.

**Zhu Tian-Yu**, born in 1993, M.S. candidate. Her research interests include Data Mining and Recommender System.

**Huang Zhen-YA**, born in 1992, M.S. candidate. His research interests include Educational Data Mining, Recommender System.



**Chen En-Hong**, born in 1968, Ph. D. professor. His research interests include Machine Learning, Data Mining, Social Network and Personalized Recommender System.

**Liu Qi**, born in 1986, Ph. D. associate professor. His research interests include Data Mining and Knowledge Discovery in Database, Machine Learning Method and Application.

**Wu Run-Ze**, born in 1990, Ph. D. candidate. His research interests include Educational Data Mining and Cognitive Diagnosis.

**Wu Le**, born in 1988, Ph. D. Her research interests include Educational Data Mining and Knowledge Discovery in Database, Recommender System, Social Network.

**Su Yu**, born in 1984, Ph. D. candidate. His research interests include Data Mining, Image Identification.

**Chen Zhi-Gang**, born in 1977, Ph. D. candidate. Junior professional titles. His research interests include Nature Language Understanding, Human-Computer Interaction, Data Mining.

**Hu Guo-Ping**, born in 1977, Ph. D. engineer. His research interests include Intelligent Voice, Language Core Technology.

## Background

This paper focus on personalized question recommendation for students, which is a sub-question of Educational Data Mining (EDM). EDM describes a research field concerned by computer science, educational psychology and statistics, in which methods and techniques for exploring data originating from various educational information systems have been developed. Recently, EDM has drawn a lot of research attention e.g., on learning management system (LMS), intelligent tutoring systems (ITS) and massive open on-line course (MOOC). Based on the huge amounts of data collected from these interactive learning environments, lots of meaningful researches have been published, e.g., in the International Conference on Educational Data Mining (EDM), International Conference on Intelligent Tutoring System (ITS) and ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Most of existing research fall into three categories: First, modeling student through data mining method or cognitive diagnosis model in order to predict student performance; Second, grouping students through clustering based on homogeneous or heterogeneous theory to explore collaborative learning patterns; Third, mining frequent patterns from the interactive feedback between students and the learning systems so as to recommend personalized remedy to students.

In our work, we combined the first and the third aspects to do a personalized question recommendation: first, we combined Probabilistic Matrix Factorization (PMF), one of the data mining method, and DINA model, one of the cognitive diagnosis model, to predict student performance (PSP); second, we recommend questions to each student by their predicted performance. In this way, we can recommend useful questions for each student through massive questions.

This research was partially supported by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), the Natural Science Foundation of China (Grant No. 61403358) and the Anhui Provincial Natural Science Foundation (Grant No. 1408085QF110). Qi Liu acknowledges the support of the Youth Innovation Promotion Association, CAS (No. 2014299).