

一种面向教育评估的智能教育辅助平台

黄振亚¹, 苏 喻², 吴润泽¹, 刘玉苹¹, 刘 淇¹, 陈志刚¹, 胡国平³

(1. 中国科学技术大学, 计算机科学与技术学院, 安徽合肥 230027;

2. 安徽大学, 计算机科学与技术学院, 安徽合肥 230039; 3. 科大讯飞股份有限公司, 安徽合肥 230088)

摘要: K-12 教育是教育领域一个重要的方面. 近年来, 在线教育模式已成为 K-12 教育重要的组成部分, 被人们广泛接受. 然而, 现有的在线教育系统和智能教育系统大多为基于教育资源数据库, 能够为学生和老师提供优质的线上教育资源, 这些针对 K-12 教育的解决方案鲜有结合线下考试(如期末考试等)资源, 并且较少考虑教师和学生的个性化需求. 针对以上问题, 提出并实现了一种面向教育评估的智能教育辅助平台(intelligent tutoring platform for educational assessment, ITPEA), 该平台采用线下考试-线上评估的模式, 从试题、学生、教师三个角度进行深入的诊断分析. 具体地, 该平台收集大规模线下考试数据, 结合教育领域模型, 首先给出对考试的试题参数分析. 其次, 针对学生知识点掌握进行分析, 给出个性化的学生认知诊断. 最后, 使用一种除去非教师因素的教师影响力分析法来评估教师, 并通过数据挖掘技术, 帮助教师找出异常学生. ITPEA 的关键技术已经在国内某在线教育系统上运行, 取得了良好的效果.

关键词: K-12 教育; 教育评估; 知识层次结构; 个性化需求; 数据挖掘

中图分类号: TP391 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2015.10.007

引用格式: HUANG Zhenya, SU Yu, WU Runze, et al. An intelligent tutoring platform for educational assessment[J]. Journal of University of Science and Technology of China, 2015, 45(10):846-854.

黄振亚, 苏 喻, 吴润泽, 等. 一种面向教育评估的智能教育辅助平台[J]. 中国科学技术大学学报, 2015, 45(10):846-854.

An intelligent tutoring platform for educational assessment

HUANG Zhenya¹, SU Yu², WU Runze¹, LIU Yuping¹, LIU Qi¹, CHEN Zhigang¹, HU Guoping³

(1. Department of Computer Science and Technology, University of Science and Technology of China, Hefei, 230027;

2. Department of Computer Science and Technology, Anhui University, Hefei, 230039;

3. USTC iFLYTEK Co., Ltd., Hefei 230088)

Abstract: K-12 education is an important part of educational psychology. Recently, online-learning has been widely accepted because of its significant effect on K-12 education. However, mostly based on the educational database, though existing online-learning systems and intelligent tutoring systems can provide useful resources for teachers and students, they seldom utilize offline test data to offer personalized services. To the end, an intelligent tutoring platform for educational assessment(ITPEA) was proposed and implemented. The platform combines both offline examinations and online resources to offer analysis

收稿日期:2015-08-27; 修回日期:2015-09-29

基金项目:国家自然科学基金(61403358), 安徽省自然科学基金(1408085QF110), 中央高校基本科研业务费专项基金(wk2350000001, wk011000000042)资助.

作者简介:黄振亚,男,1992年生,硕士生,研究方向:数据挖掘. E-mail: huangzhy@mail.ustc.edu.cn

通讯作者:刘淇,男,博士/副研究员. E-mail: qiliuql@ustc.edu.cn

from the perspectives of test papers, students and teachers. Specifically, based on the data from offline examinations, educational theories were first employed to evaluate the quality of the questions. Then, diagnostic models were constructed on the key knowledge points that students should master to meet one's personalized demands. Finally, new analytical methods for evaluating teacher's influence on class abilities were proposed with data mining technologies to help find "unusual" students in class. The core technology of ITPEA is currently operating on an online-learning system, and obtains good results.

Key words: K-12 education; educational assessment; knowledge points; personalized demands; data mining

0 引言

随着互联网技术的飞速发展,教育信息化步伐逐步加快,教育模式呈现出多样化趋势,智能教育引起了工业界和国内外学者的广泛关注^[1-2]。一方面,在工业界,各类智能教育系统层出不穷,它们通过在线网站为教师和学生提供教育服务^[3],如 MOOC 平台等。另一方面,在国内外的研究中,针对智能教育系统积累的教育数据,教育数据评估也成为重要的研究问题,结合统计学、数据挖掘、机器学习等相关技术,教育数据挖掘(educational data mining, EDM)也成为了一个独立的研究领域^[4-5]。

目前,针对教育模式不同,智能教育系统及其相关教育评估分析可大致分为以下三类^[4,6]: ① 离线教育,主要以 K-12 教育(学生 6 岁-18 岁的教育)为主。教师和学生主要参与以面对面的学校教育,运用统计学和心理学等相关技术进行教育评估。② 教育管理系统(learning management system, LMS),它通过使用 Web 挖掘和数据库技术提供多样性的优质在线教育服务。③ 智能教辅系统(intelligent tutoring system, ITS),除了提供在线教育资源外,它也使用一定的机器学习和数据挖掘技术及时给在线教师和学生一定的教育评估反馈。

近年来,LMS 和 ITS 提供的便利教学服务着实影响着 K-12 教育。国内外已有教育机构尝试结合互联网资源改善 K-12 离线教育模式,通过计算机技术辅助进行教育评估,取得了一定的成果^[3,6]。

由于 K-12 教育的特点(学生年龄小,自主学习能力相对较弱等),线下教学考试-线上教育评估的模式是 K-12 教育应用广泛的模式。现有的在线教育系统鲜有结合学生大规模线下考试数据的,使得对 K-12 教育的分析评估不够完整。

通过分析线下考试数据,可以得到阶段性的可靠的教育认知分析。目前,K-12 教育线下考试数据的教育评估任务中最重要的三类对象是试题、教师

和学生,准确地对其进行建模分析也是多种教育应用的基础,如预测学生得分、学生试题推荐、改善教师教学计划等,具有极大的价值^[7]。针对试题评估分析,学者们从难度、区分度等方面修正试题分值以及从信度、效度等方面评估试题质量。目前,应用最为广泛的是经典测试理论(classic test theory, CTT)^[8],它把线下考试得分看作真分数和误差分数的线性组合,通过一定的统计方法给出试题分析结果。针对学生诊断分析,传统统计学方法可以从均分、得分率、排名角度建立学生分析理论,但统计学方法局限在试题粒度上,没有对学生深层次的知识技能进行诊断分析,不具有普遍性(学生在试题上的表现受试题参数影响很大,而在知识点上的表现相对稳定)^[9-10]。针对此问题,一方面,教育领域中的学者们提出了多种认知诊断模型对学生能力进行建模。其中,项目反应理论(item response theory, IRT)^[11]和 deterministic inputs, noisy "and" gate model (DINA)模型^[12]是应用最广的两类模型。IRT 模型引入试题参数(如难度,区分度,猜测系数),结合学生的试题作答情况,进行学生建模,诊断每一个学生的潜在能力。DINA 模型通过引入 Q 矩阵(试题知识点关联矩阵,描述每一道试题所考察的知识点明细)^[13],将学生潜在知识能力描述为一个多维知识点掌握向量,以保证教育分析结果的可解释性。虽然 DINA 模型已经可以深入诊断学生知识点层次上的能力掌握,但它只能给出离散型的分析结果(知识点掌握或知识点未掌握),其信息量有限,难以满足学生的个性化需求。另一方面,随着数据挖掘技术的广泛运用,矩阵分解也被用于进行学生的诊断建模^[14-16],且在学生成绩预测任务中取得了良好的效果,但由于学生潜在能力向量维度的含义无法明确说明^[15,17],导致其无法给出可解释的诊断分析。针对教师诊断分析,目前最为常用的是通过其指导的学生分析结果来评估教师的能力。对于线下考试数据,常用的方法是通过分析班级平均分、班级学生排

名或基于 CTT 分析得到的统计指标对教师进行认知诊断,但是这种方法无法消除非教师因素(班级间、学生间的个体能力强弱),难以诊断真实的教师贡献程度。例如,一个平均能力强的班级在所有的统计指标中的结果都比一个平均能力弱的班级高,但这不能反映教师对班级的影响力贡献水平。目前的教育系统中评估方法仍存在以下两个主要的问题:①对学生而言,如何能够针对学生知识点掌握水平,给出学生更深入的诊断分析,以满足学生个性化需求?②对教师而言,如何合理地消除非教师因素而给出教师教学影响力的分析报告?

为了解决上述问题,本文提出并实现了一种面向教育评估的智能教育辅助平台(ITPEA)。首先,针对学生线下考试数据,结合经典测试理论(CTT)给出试题参数分析,以帮助学生、教师等理解试题。其次,结合认知诊断模型并得到学生在知识点上的掌握程度,深入挖掘学生的认知属性(失误率、猜对率等),帮助学生掌握自身能力情况,提高学习效率。最后,结合试题和学生的分析,本文提出一种除去非教师因素的教师影响力分析方法,并结合数据挖掘技术帮助教师理解班级学生,改善教学。本文的主要贡献总结如下:

(I)本文提出并实现了一个面向教育评估的智能教育辅助平台(ITPEA),整合线下考试数据,深入评估教师和学生的能力,以帮助其改善教育教学。

(II)本文结合教育心理学领域的思想,分析学生能力(知识点掌握程度、猜测率、失误率等),提供学生的个性化认知诊断结果,帮助学生提高学习效率。

(III)本文基于班级学生知识点掌握,提出一种消除非教师因素的教师分析方法,并结合数据挖掘技术,找出教师教学中的异常学生,改善教学。

1 相关工作

本节将从教育评估(试题评估、学生诊断、教师分析)和教育系统两个方面总结针对教育评估分析的相关技术和产品。

1.1 试题评估

试题评估是教育心理学的重要问题。对于一次标准的线下测试,学者对于其试题的评估方法有两类:一类是基于统计的评估方法,它通过分析试题得分的均分、方差等因素评价一道试题的难度和稳定性。另一类是经典测量理论^[8],它将得分看作真分数

和误差分数的线性组合,通过学生大量的答题记录从多个角度衡量试题质量(难度、区分度、信度、效度等),具有广泛的应用前景。

1.2 学生诊断

为了给出更精准的学生分析诊断,近年来,学生建模任务已经引起了来自教育领域和数据挖掘领域两方面学者的关注。一方面,教育心理学家提出了多种认知诊断模型(cognitive diagnosis model, CDM)^[10,18],以项目反应理论(IRT)和 DINA 模型应用最为广泛。其中,项目反应理论(IRT)为代表^[11]将学生描述成一个一维的潜在能力,可以被看作是学生的综合能力值,并通过引入试题参数(如难度、区分度、猜测系数),拟合大量答题记录,得到分析结果。IRT 模型在学生得分预测等应用中取得了良好的效果,但由于其对学生的建模是一维的潜在能力值,可描述的信息有限。DINA 模型^[12]通过引入 Q 矩阵,将学生能力描述成一个多维知识点掌握向量,该向量反映学生是否掌握试题所考查的每一个知识点,从知识点的角度刻画学生能力来获得可解释的诊断结果。然而,DINA 模型描述的是学生知识点的离散掌握度,不能精准地诊断学生在主观题上的能力水平,为此有学者结合模糊系统,提出一种针对主客观题通用的学生建模方法^[19]。另一方面,矩阵分解技术也被广泛用于学生建模^[14-16],它通过分解学生考试数据构造学生和试题低维矩阵来刻画学生潜在能力,并通过这种方式在学生得分预测和学习资源推荐上取得了良好的效果,但由于学生和试题的低维的潜在能力维度的含义难以明确说明,它在教育评估的发展中受到了限制^[15,17]。

1.3 教师评价

学者对于线下考试的教师评价任务主要是通过班级或学生表现情况来进行教育评估的^[20-21]。现有的教师评价方法主要分为如下几种^[22]:基于统计的比较评价和基于学生成绩增值的评价。其中,基于统计的比较评价是目前 K-12 教育使用比较广泛的教师评价手段^[23],它通过统计学生参数(如平均得分、平均排名等)和经典测量理论(CTT)得到的相关参数(如班级得分率、班级通过率等)来评价教师,其方法简单并易于理解,但是无法消除来自学生和班级个体自身因素的影响,使评价结果的可信度降低。基于增值的教师评价方法通过比较相关学生两次测试的成绩增值来评估教师^[24],这种方法在一定程度上消除了学生间的个体差异,但其分析仍处在试题粒

度上,而学生对于试题的表现易受到试题参数影响(难度等),没有从相对稳定的知识点技能层面进行分析,使得分析结果的可靠性存疑.此外,随着数据挖掘技术的广泛应用,学者采用聚类分类算法进行学生聚类分析^[25-26],从而进一步评价教师能力水平,取得了一定的效果.

1.4 教育系统

针对 K-12 教育的教育辅助产品在国内外均有一定的尝试.在国内,网上阅卷模式是运用最广的模式,如全通教育^[27]等.它们采用网上阅卷的方法节省教学时间,结合一定的统计指标进行简单的教师学生分析,给予教学反馈.另外,也有教育系统通过组建在线题库帮助学生,如猿题库^[28]等,它们根据学生在线做题的信息,通过简单的 LR 方法对学生建模,预测学生成绩.在国外,Knewton^[29]提出一种自适应学习模式,通过 IRT, PGM 等进行教育评估并给出推荐.

2 平台框架介绍

本节将介绍面向教育评估的智能教育辅助平台(ITPEA)的系统框架图,如图 1 所示.该平台主要评估 K-12 教育中试题、学生和教师三类对象,自上而下可以分成四个部分:①.数据收集层.②.数据建模层.③.数据分析层.④.数据展示层.具体地,在数据收集层,ITPEA 收集标准线下考试数据,一方面扫描学生考试试卷,组织教师网上阅卷;另一方面标注

试题的知识点关联信息(即试题与知识点、抽象能力的关联信息),然后将信息统一录入后台教育资源数据库,试题 Q 矩阵信息的生成可以保证该平台对试题、教师、学生诊断分析结果具有很好的解释性^[13].在数据建模层,该平台结合教育心理学中经典测试理论(CTT)和多种认知诊断模型(IRT, DINA)对含有标注知识点的线下考试数据进行建模,从试题参数、学生属性等多个角度得到多个知识点层面的信息.在数据分析层,该平台由试题评估模块、学生诊断模块和教师分析模块组成.在试题评估模块,该平台给出试题的难度、区分度、信度等指标,帮助教师和学生理解试题,评估试题质量.在学生诊断模块,该平台深入挖掘学生的能力结构、知识点层次及其自身的认知属性(失误率、猜测率),以实现对学生个性化诊断评估,并结合相关诊断结果,给学生一定的学习建议.在教师分析方面,该平台结合试题参数和相关学生的诊断结果,给出教师影响力分析(除去来自班级学生自身的非教师因素),并运用数据挖掘技术(如聚类分析、异常点挖掘等),给出如在班级等粒度下的班级整体分析报表,为教师制定个性化教学提供依据.在数据展示层,教师和学生可以在线查询自己的分析报表,使得教师和学生能快速评估试题;使得学生迅速掌握自身技能水平,查漏补缺,提高学习效率;使得教师准确理解班级学生,改善教学.

3 平台实现和技术细节

本节将具体介绍 ITPEA 从数据建模层到数据分析层的各模块详细的技术方案.经过数据收集层,针对学生答题情况,下面首先介绍如何利用 CTT 评估试题参数.然后介绍如何进行学生画像诊断,得到学生知识点层面的信息.最后介绍如何结合试题和学生诊断进行教师评价.

3.1 试题评估

试题评估是教育研究的重要问题,有效的试题评估方法可以帮助教师和学生寻找合适的教育资源,节省教育教学时间,提高教师教学和学生学习的效率.ITPEA 结合经典测试理论(CTT)^[8-9],从测试难度、区分度、信度等角度评估试题质量.技术路线流程图如图 2 所示.

图 2 试题评估细节如下:

难度:试题难度表示一道试题的难易程度.试题的难度可以决定一次测试的得分分布.一般地,试题

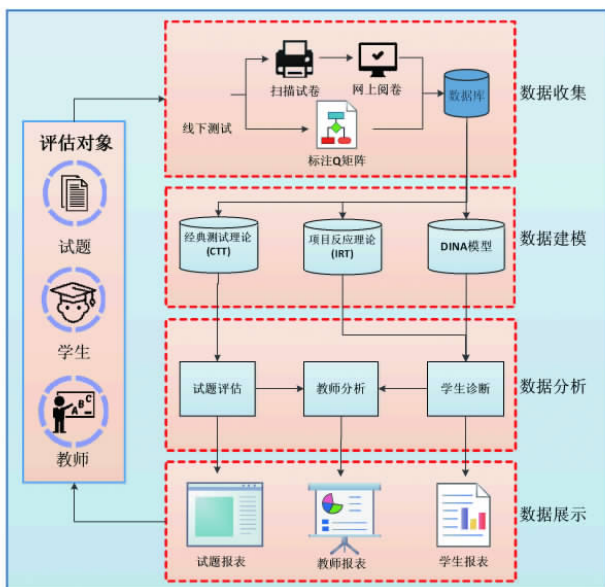


图 1 教育评估的智能教育辅助平台框架

Fig. 1 Intelligent tutoring platform for educational assessment

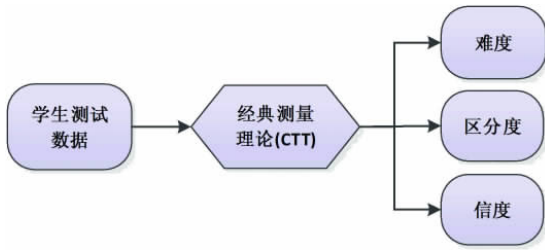


图 2 试题评估技术流程

Fig. 2 The flowchart of test evaluation

i 的难度 a_i 可以由试题的平均得分率衡量.

$$a_i = 1 - \frac{\bar{X}_i}{Y_i}, a_i \in (0, 1) \quad (1)$$

式中, $\bar{X}_i = \frac{1}{J} \sum_{j=1}^J X_{ji}$ 表示试题 i 上所有学生的平均得分, Y_i 表示试题 i 的总分.

区分度: 试题区分度表示一道试题对于学生的区分能力, 具有良好区分度的试题, 能够将不同水平的学生区分开. 即在具有良好区分度的试题上, 能力强的学生能够得高分, 而能力弱的学生只能得到低分. 根据教育心理学理论, 试题 i 区分度 b_i 的计算由试题高分组学生 (27%) 和试题低分组学生 (27%) 的通过率的差值.

$$b_i = p_H - p_L, b_i \in (-1, 1) \quad (2)$$

式中, p_H, p_L 表示试题 i 上高分组学生和低分组学生的得分率 (通过率).

一次测试的难度 A 和区分度 B 可以由试题上难度和区分度的均值来评估, 即

$$A = \frac{1}{Y} \sum_{i=1}^I Y_i \times a_i \quad (3)$$

$$B = \frac{1}{Y} \sum_{i=1}^I Y_i \times b_i \quad (4)$$

式中, Y 表示测试总分, Y_i, a_i, b_i 表示试题 i 的总分、难度和区分度.

测试信度: 测试信度表示测试结果的稳定性和可靠性. 一次良好的测试应该具有较高的信度, 以保证测试和评估结果的可靠. 测试信度 R 常用克隆巴赫 α 系数衡量.

$$R = \frac{I}{I-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma^2} \right) \quad (5)$$

式中, I 表示试题总数, σ_i^2 表示试题 i 的方差, σ^2 表示学生试题得分总分的方差.

3.2 学生诊断

学生诊断 (学生画像) 的目标是挖掘学生在知识层面的技能掌握情况, 分析学生在知识点层面的

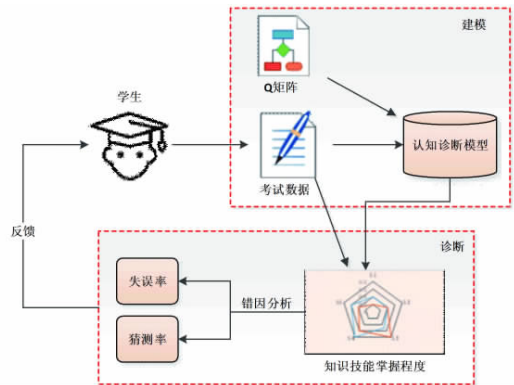


图 3 学生诊断技术流程

Fig. 3 The flowchart of student diagnostics

表现比学生在试题层面的表现有更可靠的结果. 学生诊断可以帮助学生发现自身的薄弱知识点, 从而有针对性地进行学习, 提高效率. 为了得到具有解释性的学生个性化诊断结果, 如图 3 所示. ITPEA 首先结合认知诊断模型, 在知识点层面对学生进行建模, 得到学生知识点掌握程度, 然后进一步分析学生的认知属性 (失误差、猜对率).

知识点掌握程度: DINA 模型是一种离散型的认知诊断模型^[10,12], 它将学生的能力描述为一个多维知识点技能向量 $\alpha_j = (\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jk})$, 且定义试题 i 关联一组知识点技能 $q_i = (q_{i1}, q_{i2}, \dots, q_{ik})$. 则学生 j 对于试题 i 的掌握情况定义 δ_{ji} 为:

$$\delta_{ji} = \prod_{k=1}^K \alpha_{jk}^{q_{ik}} \quad (6)$$

式中, K 表示测试考察的知识点技能总数; α_{jk} 表示学生 j 在技能 k 上的掌握情况 ($\alpha_{jk} = 1$ 表示学生 j 掌握技能 k , $\alpha_{jk} = 0$ 表示学生 j 未掌握技能 k); q_{ik} 表示学生 i 是否考察技能 k ($q_{ik} = 1$ 表示试题考察技能 k , $q_{ik} = 0$ 表示试题 i 未考察技能 k), 因此 δ_{ji} 表明学生 j 只有掌握试题 i 所考察的所有技能, 才能正确作答试题 i .

此外, DINA 模型引入两个试题上的参数 (即失误差和猜测率) 来诊断学生的答题情况. 具体地, 学生 j 在试题 i 上的真实答题情况为:

$$P(X_{ji} = 1 | \alpha_j, s_i, g_i) = g_i^{1-\delta_{ji}} (1 - s_i)^{\delta_{ji}} \quad (7)$$

式中, g_i 为试题 i 上的猜测率, 表示学生在未掌握试题 i 考查的所有知识点的情况下仍可答对试题 i 的概率; s_i 为试题 i 上的失误差, 表明学生掌握试题 i 考查的所有知识点的情况下未能答对试题 i 的概率.

DINA 模型只能得到学生 j 在技能 k 上离散的掌握情况 (即掌握和未掌握), 而在实际情况中, 学生

对于知识点的掌握程度是一个连续的值,学生对知识点的掌握程度有高有低,因此 ITPEA 进一步处理 DINA 模型的结果,通过计算学生技能向量 α 的后验概率来获得学生知识点掌握程度.具体地,学生 j 在技能 k 上的掌握程度 α_{jk}' 可以表示为:

$$\alpha_{jk}' = P(\alpha_{jk} = 1 | X_j) = \frac{\sum_{\alpha_{jk}=1} P(\alpha_x | X_j)}{\sum_{x=1}^{2K} P(\alpha_x | X_j)} \quad (8)$$

根据学生知识点掌握程度,ITPEA 针对学生,通过考虑学生答题情况的后验概率分析学生在试题上掌握程度.具体地,学生 j 在试题 i 上的掌握程度 X_{ji}' 表示为:

$$\left. \begin{aligned} P(X_{ji}' | X_{ji} = 1) &= \frac{M_{ji} \times (1 - s_i)}{M_{ji} \times (1 - s_i) + (1 - M_{ji}) \times g_i} \\ P(X_{ji}' | X_{ji} = 0) &= \frac{M_{ji} \times s_i}{M_{ji} \times s_i + (1 - M_{ji}) \times (1 - g_i)} \end{aligned} \right\} \quad (9)$$

式中, $M_{ji} = \sqrt[\sum_{k=1}^{q_{ik}}]{\prod_{q_{ik}=1} \alpha_{jk}'}$ 表示在试题 i 关联的所有技能 k 上,学生 j 的知识点掌握程度的几何平均值.

学生属性分析:进一步,ITPEA 可以分析出学生在试题上的错因,给出学生层面上的猜对率和失误率.具体地,学生 j 在试题 i 上的猜对率 g_{ji} 和失误率 s_{ji} 分别为:

$$\left. \begin{aligned} P(g_{ji} | X_{ji} = 1) &= \frac{X_{ji}' \times s_i}{X_{ji}' \times s_i + (1 - X_{ji}') \times (1 - g_i)} \\ P(s_{ji} | X_{ji} = 0) &= \frac{(1 - X_{ji}') \times g_i}{X_{ji}' \times (1 - s_i) + (1 - X_{ji}') \times g_i} \end{aligned} \right\} \quad (10)$$

由于学生在试题上的表现受到试题因素影响较大,在知识点上的表现相对稳定(如考查相同知识点的难易两道题,学生在这两道题上的表现可能差异很大).相比于针对试题层面的分析,ITPEA 的学生诊断着眼于学生知识点层面上,这种方法的学生诊断结果具有较高的可靠性.同时,结合教育心理学 Q 矩阵,保证了诊断结果的可解释性.学生通过自身知识点技能水平、失误率和猜测率,可以及时查漏补缺,做到有针对性的学习.

3.3 教师分析

如图 4 所示,结合试题评估和学生诊断的结果,ITPEA 提出一种除去非教师因素(班级学生个体因素)的教师分析方法.首先给出教师在班级和知识点

层面的影响力分析.同时,结合聚类 and 异常点检测等数据挖掘技术,帮助教师发现班内的异常学生,使得教师的评价更准确,以帮助教师改善教学.

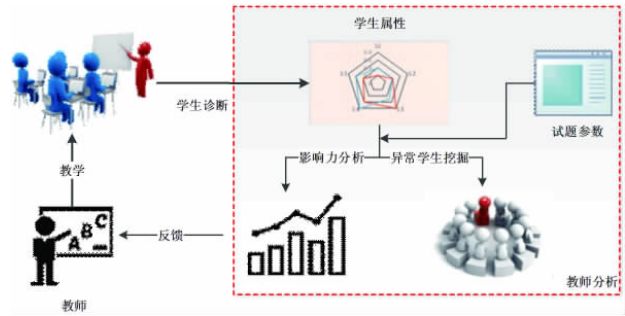


图 4 教师分析技术路线流程

Fig. 4 The route of teacher analysis technology

教师影响力分析:在 K-12 教育中,针对教师的分析主要由教师教授的学生能力变化所体现,一般以班级为粒度进行相关分析,且教师分析是一个连续的过程,根据增量式的理论,教师对班级学生的影响力可通过 t 时刻班级知识点技能的掌握程度与 $t-1$ 时刻班级知识点技能掌握程度的差值来体现,因为学生在知识点技能层次上的表现相对于试题和测试粒度是稳定的,因此时刻 t 班级内某学生 j 在知识点技能 k 上的得分率 p_{jk}^t 为:

$$p_{jk}^t = \frac{\sum_{i=1}^I \frac{X_{ji}}{Y_i} \times \alpha_{jk} \times q_{ik}}{\sum_{i=1}^I q_{ik}} \quad (11)$$

式中, X_{ji} 分别表示学生 j 在试题 i 上的得分, α_{jk} 表示学生 j 在知识点技能 k 上的掌握程度, Y_i 表示试题 i 上的总分.

进一步,针对班级内的所有学生,可以得到时刻 t 班级 n 在知识点技能 k 上的平均得分率 p_{nk}^t 为:

$$p_{nk}^t = \frac{\sum_{j=1}^J p_{jk}^t}{J} \quad (12)$$

式中, J 表示班级 n 中的学生总数.

对于时刻 t 的测试,教师 u 对于班级 n 在知识点技能 k 上的影响力 In_{nk}^t 可表示为:

$$In_{nk}^t = \frac{p_{nk}^t - p_{nk}^{(t-1)}}{p_{nk}^{(t-1)}} \quad (13)$$

式中, $p_{nk}^t, p_{nk}^{(t-1)}$ 表示 t 时刻、 $t-1$ 时刻班级 n 在知识点技能 k 上的平均得分率.

进一步,对于时刻 t 的测试,教师 u 对于班级 n 的影响力 In_n^t 为:

$$In_n^t = \frac{1}{K} \sum_{k=1}^K In_{nk}^t \quad (14)$$

式中, K 表示时刻 t 的测试所考察的技能总数.

异常学生挖掘: 对于 K-12 教育而言, 帮助教师迅速准确找到班级内的“异常学生群体”(明显退步, 能力与班级能力有明显差距)是有重要意义的. 它能够帮助教师有针对性地指导学生学习, 提高效率. 具体地, ITPEA 使用班级内学生知识点的掌握程度作为学生特征进行聚类分析^[25-26], 改进了现有方法使用学生得分等试题上的参数, 找出班级知识点掌握相似的学生簇, 进而分析出异常学生群体.

4 平台展示

面向教育评估的智能教育辅助平台的关键技术



图 5 试题分析

Fig. 5 Papers analysis

已经在国内某在线教育系统上运行, 覆盖全国 200 多个学校, 在 K-12 教育中取得了良好的效果. 本节将具体展示该平台的试题分析、学生报表和教师报表. 最后, 本文将通过平台在相关试点学校的分析反馈样例, 说明分析结果的有效性.

4.1 平台结果展示

试题分析: 图 5 展示 ITPEA 针对试题的分析结果, 它展示了试题与其考查知识点的关联情况 (a), 如第 3 题考查等比数列和函数与方程两个知识点; 并进一步通过试题, 分析试卷参数 (b) (难度、区分度等).

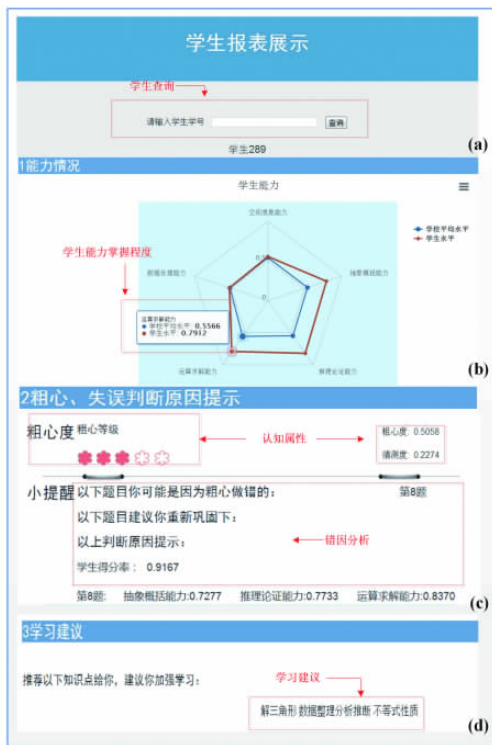


图 6 学生报表

Fig. 6 The student report

学生报表: 图 6 以学生报表的形式展示了 ITPEA 的学生模块的相关分析结果. 自上而下, 学生报表可分为 4 个部分. 对于 (a) 中一个查询的特定学生 (图 6 中查询学生 289), ITPEA 可以根据学生本次线下考试的测试结果, 分析出学生在本次考试中知识点技能的掌握程度, 并与全校情况作对比 (b), 以帮助学生了解自己的知识层次水平. 同时, 该学生还可以看到本次测试自身的认知属性 (失误率, 猜对率等) 以及相关典型题目上的“错因分析” (c) (如图 6, 该学生第 8 题可能是粗心做错的). 进一步, ITPEA 根据该学生本次的整体知识点情况, 给出学生知识点上的学习建议 (d), 以帮助学生查漏补缺, 做到自身个性化有针对性的学习, 提高学习效率.

教师报表: 图 7 以教师报表的形式展示了 ITPEA 的教师模块的相关分析结果. 自上而下, 教师报表可分为 5 个部分. 对于 (a) 中一个查询的特定教师 (图 7 中查询 C 中学 1 班数学科目), ITPEA 可以根据班级学生线下考试结果, 分析出教师对所教授的班级的影响力 (b), 同时针对线下考试所涉及的知识技能, ITPEA 进一步给出教师对班级各知

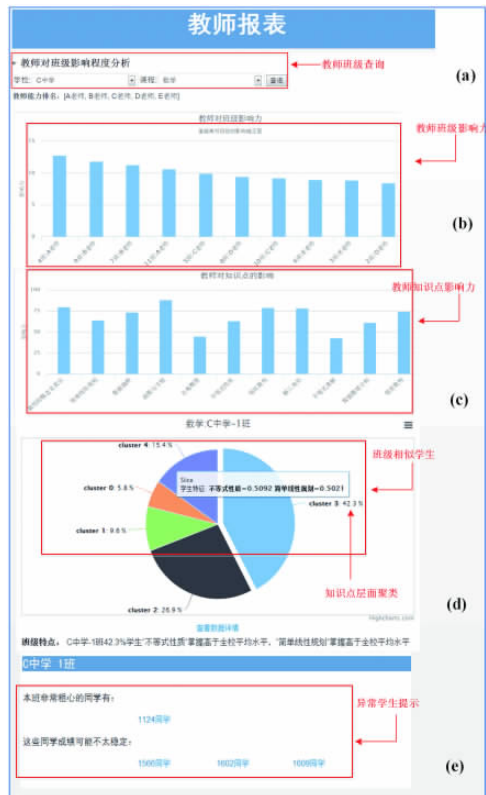


图 7 教师报表

Fig. 7 The teacher report

识点的影响力(c),以帮助教师快速理解班级能力水平,改善教学.此外,根据班级学生知识点技能掌握水平,使用数据挖掘相关技术,ITPEA 将班级学生从知识点层次上进行聚类划分(d),具有良好的解释性,并通过异常学生发现,ITPEA 可以帮助教师找到班级内需要关注的学生(e)(图 7 中展示了 C 中学 1 班本次数学科目考试中容易失误的学生和成绩不稳定学生),以帮助教师给异常学生进行个性化的指导.

5 结论

本文针对 K-12 线下教育,提出并实现一个面向教育评估的智能教育辅助平台(ITPEA),该平台收集整合 K-12 线下考试资源,从试题、学生、教师三个角度进行教育评估,以帮助教师学生评估试题质量;帮助学生查漏补缺,提高学习效率;帮助老师掌握班级学生的能力水平,做到有针对性地教学,改善教学质量.具体地,针对试题,ITPEA 结合经典测试理论(CTT)给出试题和测试粒度上的难度、区分度、信度、效度等指标,评估试题和测试的质量.针对学生,ITPEA 克服传统学生评估方法受到试题参数

影响导致评估结果可信度不足的问题,结合认知诊断模型,从学生知识点技能层面对学生进行诊断分析,并进一步给出了学生的认知属性分析,帮助学生查漏补缺,做到自身个性化有针对性地学习,提高学习效率.最后针对教师,结合学生知识点技能掌握情况,ITPEA 从班级知识点角度通过分析非教师因素影响,给出了教师分析结果,并帮助教师找到班内异常学生,做到个性化的指导,改善教学质量.

为了提高 K-12 教育评估的可靠性、精确性,未来研究工作可以从以下几个方面进行:①针对试题,进一步考虑试题与知识点技能之间的相关性,以实现试题属性自动标注.②针对学生,进一步考虑学生在学习过程中知识点技能的变化情况,以设计时序化的学生知识点模型,更好地对学生建模.③针对教师,进一步考虑教师教学特点,以精确建模教师教学影响力,评估教师能力.

参考文献(References)

[1] Cha H J, Kim Y S, Park S H, et al. Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system [C]// Proceedings of the 8th International Conference on Intelligent tutoring systems. Taiwan: Springer, 2006: 513-524.

[2] Burns H, Luckhardt C A, Parlett J W, et al. Intelligent Tutoring Systems: Evolutions in Design [M]. Psychology Press, 2014.

[3] Anderson A, Huttenlocher D, Kleinberg J, et al. Engaging with massive online courses [C]// Proceedings of the 23rd International Conference on World Wide Web. Seoul, Korea: ACM Press, 2014: 687-698.

[4] Romero C, Ventura S. Educational data mining: a review of the state of the art [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2010, 40(6): 601-618.

[5] Scheuer O, McLaren B M. Educational data mining [C]// Encyclopedia of the Sciences of Learning. New York, USA: Springer, 2012: 1075-1079.

[6] Baker R S J D, Yacef K. The state of educational data mining in 2009: A review and future visions [J]. JEDM-Journal of Educational Data Mining, 2009, 1 (1): 3-17.

[7] Calders T, Pechenizkiy M. Introduction to the special section on educational data mining [J]. ACM SIGKDD Explorations Newsletter, 2012, 13(2): 3-6.

[8] DeVellis R F. Classical test theory [J]. Medical Care,

- 2006, 44(11S): S50-S59.
- [9] Fan X T. Item response theory and classical test theory: An empirical comparison of their item/person statistics [J]. Educational and Psychological Measurement, 1998, 58(3): 357-381.
- [10] DiBello L V, Roussos L A, Stout W. 31A Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models [M]. Handbook of statistics, 2006, 26: 979-1030.
- [11] Harwell M R, Baker F B, Zwarts M. Item parameter estimation via marginal maximum likelihood and an EM algorithm: A didactic [J]. Journal of Educational and Behavioral Statistics, 1988, 13(3): 243-271.
- [12] de La Torre J. DINA model and parameter estimation: A didactic [J]. Journal of Educational and Behavioral Statistics, 2009, 34(1): 115-130.
- [13] Rupp A A, Templin J. The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model [J]. Educational and Psychological Measurement, 2008, 68(1): 78-96.
- [14] Thai-Nghe N, Drumond L, Horváth T, et al. Factorization techniques for predicting student performance [J]. Educational Recommender Systems and Technologies: Practices and Challenges, 2011, 37(2): 157-186.
- [15] Mnih A, Salakhutdinov R. Probabilistic matrix factorization [C]//Advances in neural information processing systems, 2007: 1257-1264.
- [16] Castro F, Vellido A, Nebot À, et al. Applying data mining techniques to e-learning problems [A]//Evolution of Teaching and Learning Paradigms in Intelligent Environment. Berlin Heidelberg: Springer, 2007: 183-221.
- [17] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems [J]. Computer, 2009, 42(8): 30-37.
- [18] Nichols P D, Chipman S F, Brennan R L, et al. Cognitively Diagnostic Assessment [M]. Mahwah, USA: Lawrence Erlbaum Associates, 1995.
- [19] Wu R Z, Liu Q, Liu Y P, et al. Cognitive modelling for predicting examinee performance [C]//International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina: ACM Press, 2015.
- [20] Teven J J, McCroskey J C. The relationship of perceived teacher caring with student learning and teacher evaluation [J]. Communication Education, 1997, 46(1): 1-9.
- [21] Baker E L, Barton P E, Darling-Hammond L, et al. Problems with the use of student test scores to evaluate teachers [R]. EPI Briefing Paper # 278, Economic Policy Institute, 2010.
- [22] Darling-Hammond L, Beardsley A, Haertel E, et al. Evaluating teacher evaluation: What we know about value-added models and other methods [J]. Phi Delta Kappan, 2012, 93(6): 8-15.
- [23] Ballou D, Sanders W, Wright P. Controlling for student background in value-added assessment of teachers [J]. Journal of Educational and Behavioral Statistics, 2004, 29(1): 37-65.
- [24] 任玉丹, 边玉芳. 美国学校增值性评价模式研究 [J]. 比较教育研究, 2012, (2): 76-79.
Ren Y D, Bian Y F. Study on the value added assessment System for school in America [J]. Comparative Education Review, 2012, (2): 76-79.
- [25] Bovo A, Sanchez S, Héguy O, et al. Analysis of students clustering results based on Moodle log data [C]//6th International Conference on Educational Data Mining-EDM. Memphis, USA: ACM Press, 2013: 306-307.
- [26] Shi N Y, Chen K, Li C H. The application of fuzzy clustering in teacher-evaluating model [C]//IEEE International Symposium on IT in Medicine & Education. IEEE Press, 2009, 1: 872-875.
- [27] 全通教育 [EB/OL]. <http://www.qtone.cn/>.
- [28] 猿题库 [EB/OL]. <http://www.yuantiku.com/>.
- [29] Knewton [EB/OL]. <https://www.knewton.com/>.