

文章编号: 2096-1472(2017)-12-06-03

基于项目反应理论的组卷算法研究

单美静, 刘 琴

(华东政法大学信息科学与技术系, 上海 201620)

摘 要:传统的网络考试组卷算法由于仅仅考虑考试范围,以及试卷整体难度系数,而从不考虑应试者的能力水平,从而造成生成的部分试卷无法有效地达到测试应试者能力水平的效果。项目反应理论作为一种现代心理测试理论,在评估应试者能力水平方面具有广泛的应用。本文基于项目反应理论设计了一种能充分考虑应试者能力水平的组卷算法,从而使得组成的试卷能够达到命题老师的预期效果。实际应用结果也表明,该算法是非常有效的。

关键词:项目反应理论;组卷算法;试题库建设

中图分类号: TP391 **文献标识码:** A

Research on the Test Paper Generating Algorithm Based on the Item Response Theory

SHAN Meijing, LIU Qin

(Institute of Information Science and Technology, East China University of Political Science and Law, Shanghai 201620, China)

Abstract: With the traditional network test paper generating algorithm, the scope and the overall difficulty coefficient of the test paper are the major considerations, but the ability of the examinee is neglected. Consequently, some generated test papers fail in assessing the examinee's ability. The item response theory, as a modern psychological testing theory, has been widely applied in assessing the ability of examinees. Based on the item response theory, the paper designs a test paper generating algorithm with a full consideration to the ability of examinees, so that teachers can achieve the expected effect with the generated papers. The practical application results show that the algorithm is very effective.

Keywords: item response theory; test paper generating algorithm; item bank construction

1 引言(Introduction)

试题库建设作为高校课程建设的一个必选项,具有重要的意义。首先,它是实现教改分离的重要前提,能够客观、公正地评价教师教学效果;其次,教师可以快捷地对试题进行管理和维护更新,更全面地考察学生学习的效果;最后,试题库建设还能够减轻教师的负担。然而,在很多高校,普遍存在着试题库“重建设轻维护”“重建设轻使用”的问题。究其原因,主要有以下几点:(1)很多高校老师认为试题库的组卷策略很难覆盖到教学所需要考核的知识点,在进行自动组卷后还需要进行大量的人工干预;(2)试题库一旦建设好,其题目的难度基本保持不变,但是在授课过程中可能会出现面对不同专业的学生的学习状态的情况,从而造成使用同一试题库可能出现较高不及格率的风险。这些问题的根本原因在于试题库在建设过程中,组卷时未考虑到应试者的能力水平,没有进行有效的私人定制。

项目反应理论(Item Response Theory: IRT)^[1]作为一种

现代心理测试理论,能够指导我们测试出应试者的“潜在特质”,即能力,从而能够指导我们进行测试试卷编制。项目反应理论包含有两个特性,一个是项目参数估计不变性,另一个是能力参数估计不变性。这两个特性保证了利用项目反应理论进行测试的结果,不受所选的测试样本影响,同时还能保证应试者的能力估计与被测试的项目无关^[2,3]。另外项目反应理论还能帮助命题者在试题库建设过程中,帮助命题者选择与应试者能力相匹配的题目组成试卷^[4-6]。

2 项目反应理论简介(Introduction of item response theory)

项目反应理论,有时也称作潜在特质理论或潜在特质模型,在现代心理测试领域运用非常广泛,同时也广泛应用于教育测试领域。项目反应理论的基本思想^[7]是应试者的某种潜在特质与他们对项目的反应(正确作答的概率)之间存在着一定的关系,并且这种关系可以通过数学模型表示出来,从而可以通过构建数学模型表示应试者能力、项目参数以及正

确作答的概率之间的关系。项目反应理论在计算机自适应测试中具有广泛的应用。项目反应理论是建立在“能力单维性假设”“局部独立性假设”和“项目特征曲线假设”三个基本假设基础上。项目反应理论可以根据应试者回答问题的情况，通过对测试项目特征曲线的运算来推测应试者的能力水平。项目反应理论中所建立的测试项目一般包含三个参数：区分度(Discriminative index)— a_i 、难度(Difficulty index)— b_i 和猜测系数(Guessing index)— c_i 。根据构建的数学模型所包含的参数不同，特征函数可分为仅包含难度参数的单参数模型、包含难度、区分度的双参数模型，以及包含难度、区分度和猜测系数的三参数模型。

Logistic模型作为项目反应理论中最常用的模型，是伯恩鲍姆于1957年提出的一种二级评分IRT模型，它能够很好地匹配被测试者的测试结果。与特征函数的分类类似，它也包含单参数、双参数和三参数三种模型，其表达式分别为：

$$P_i(\theta) = \frac{1}{1+e^{-D*(\theta-b_i)}} \quad (1)$$

$$P_i(\theta) = \frac{1}{1+e^{-D*a_i*(\theta-b_i)}} \quad (2)$$

$$P_i(\theta) = c_i + \frac{1-c_i}{1+e^{-D*a_i*(\theta-b_i)}} \quad (3)$$

其中， $P_i(\theta)$ 为能力为 θ 的被测试者正确作答测试项目 i 的概率；

$D=1.702$ ；

θ 为应试者能力估计值；

a_i 为测试项目 i 的区分度，其值越大，则表明被选测试项目对被测试者的区分度越强；

b_i 表示测试项目 i 的难度(难易程度)；

c_i 为测试项目 i 的猜测系数，其值越大，说明不论应试者能力高低，都容易猜对。

由于单参数、双参数Logistic模型都是三参数Logistic模型的特例，所以我们下面以三参数Logistic模型的项目特征曲线(如图1)为例，介绍计算机组卷过程中的项目特征曲线。

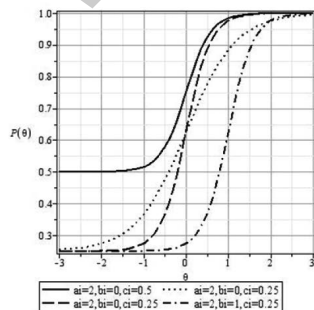


图1 不同参数的三参数Logistic模型项目特征曲线

Fig.1 Three-Parameters Logistic model item characteristic curve with different parameters

从图1的曲线可以看出，当应试者的能力值 $\theta_i = b_i$ 时，应试者正确作答的概率为 $P_i = \frac{1+c_i}{2}$ 。若不考虑猜测系数 c_i ，则正确作答和错误作答的概率皆为1/2。即对于能力值为 b_i 的应试者来说，所测试项目的难度适中。当应试者的能力值 $\theta_i > b_i$ 时， $P_i(\theta)$ 将大于0.5，并且 $P_i(\theta)$ 的值随着应试者的能力值的增大而趋近于1，即正确作答的概率越大；反之，应试者的能力值越小， a_i 越接近于0，即正确作答的概率越小。项目特征曲线的陡峭程度会随着 a_i 的值而变化， a_i 的值越大则曲线越陡峭， a_i 随应试者能力值 θ_i 的变化就越敏感，即该项目更能区分应试者的能力水平；当 b_i 增大时，项目特征曲线右移，则说明对于同一能力值的应试者，难度越高的项目越难正确作答。

教育测量学研究表明，随机选择的大规模应试者群体的基本心理素质的分布服从正态分布，所以理论上讲应试者的能力取值范围为 $(-\infty, +\infty)$ ，但在实际应用中，为了计算方便，一般取值为 $(-3, 3)$ 。

3 组卷算法分析(Analysis on the test paper generating algorithm)

有了三参数Logistic模型的理论基础，一方面我们就可以利用阶段性考试测试出所有应试者的能力水平，另一方面可以结合试题库项目的难度、知识点分布、应试者能力水平进行组卷，从而使得考试结果满足预期的正态分布以及预期目标。综上可知，基于项目反应理论的组卷算法主要包括两个部分：第一部分为应试者能力水平的测试，第二部分为基于应试者能力水平、知识点分布、预期结果的组卷算法。

算法3.1 应试者能力水平评估算法

输入：某试题库及应试者初始能力水平。

输出：某应试者的能力水平。

Step1：根据某种策略选择一个项目开始测试；

Step2：由应试者的作答情况评估应试者的能力水平；

Step3：根据新的能力水平选择一个合适的测试项目供应试者作答；

Step4：根据作答情况重新估计能力水平。如果能力水平趋于稳定(新的能力水平—旧能力水平 <0.05)，则结束；否则跳转到Step3。

算法3.2 基于项目反应理论的组卷算法

输入：试卷知识点分布、预期成绩分布。

输出：试卷测试项目。

Step1：将所有应试者按照能力水平由低到高排序；

Step2：根据考试成绩将满足正态分布的特点，按照Logistic模型选择预计80%应试者正确作答概率在0.8以上的测试项目；

Step3: 根据Logistic模型选择对于10%较高能力应试者作答概率在0.8以上的测试项目。

为了实现上述的应试者能力水平测试以及组卷过程,有如下几个问题需要解决好。(1)测试项目参数的确定;(2)初始测试项目的选择;(3)测试过程中应试者能力水平的重新评估。

3.1 测试项目参数的确定

确定测试项目的三个参数是应用项目反应理论进行组卷的前提。在实现过程中,我们采用的是应用EM算法的边缘极大似然估计法进行参数估计,以保证得出的测试项目参数具有参数不变性的特点。

3.2 初始测试项目的选择

初始测试项目的选择一般有如下几种方法:

(1)假定应试者具有中等水平的能力值,从而选择难度中等的测试项目。

(2)若系统中保存有应试者的能力水平,则根据此能力水平选择合适的测试项目。

(3)由应试者自己填写初始能力值,系统再根据此值选取合适的测试项目。

3.3 应试者能力水平的重新评估

在测试过程中,需要根据应试者作答情况对其能力水平进行评估。本文采取极大似然估计方法估计应试者能力水平。其基本过程如下:

假定某应试者对 n 个给定的客观题的作答模式为:

$$U = (U_1, U_2, \dots, U_i, \dots, U_n)$$

其中 U_i 取值1或0。若应试者正确作答了第 i 道客观题,则取1,否则取0。以 $L(U_i|\theta)$ 表示能力值为 θ 的应试者正确作答第 i 道客观题 U_i 的概率,即Logistic模型中的 $P_i(\theta)$,由局部独立性假设可计算其联合概率为:

$$L(U|\theta) = \prod_{i=1}^n P_i(\theta)^{U_i} (1 - P_i(\theta))^{(1-U_i)} \quad (4)$$

上式也称为似然函数,使(4)式达到最大值的 θ 即为 θ 的极大似然估计值 $\hat{\theta}$ 。为了计算 $\hat{\theta}$,我们对式(4)两边取自然对数,得到对数型的似然函数:

$$\ln L(U|\theta) = \sum_{i=1}^n U_i \ln P_i(\theta) + (1 - U_i) \ln(1 - P_i(\theta)) \quad (5)$$

为了计算(5)式达到最大值时的 $\hat{\theta}$,只要求解方程

$$f(\theta) = \frac{d}{d\theta} \ln L(U|\theta) = 0 \quad (6)$$

从式(6)可以看出其是非线性的,不能直接求解。此时,我们采用Newton-Raphson法进行求解。

4 结论(Conclusion)

项目反应理论在计算机自适应测试中得到了广泛应用,但是在组卷算法方面的研究还不多。本文结合项目反应理论在评估应试者能力水平方面的优势,结合考试范围以及考试成绩预期期望,设计了一种基于项目反应理论的自动组卷算法,该算法的采用能够有效地达到因材施教的目的,通过在《线性代数》《数据结构》等试题库组卷中的应用,较好地达到了预期目的。但此项工作的缺陷是测试项目只能是客观题,如何在主观题考核方面加以应用还有待进一步的研究。

参考文献(References)

- [1] Baker F.B. The basics of item response theory. Port smouth[M]. NH: Heinemann, 1985: 1-74.
- [2] Choi, Y. J., Alexeev, N., Cohen, A. S. Different item functioning analysis using a mixture 3-parameter logistic model with a covariate on the TIMSS 2007 mathematics test[J]. International Journal of Testing, 2015, 15(3): 239-253.
- [3] Frick, H., Strobl, C., Zeileis, A. Rasch mixture models for DIF detection: A comparison of old and new score specifications[J]. Educational and Psychological Measurement, 2015, 75(2): 208-234.
- [4] 刘锋, 郭维威, 等. 基于项目反应理论的计算机自适应测试算法的研究与实现[J]. 软件, 2014(7): 28-32.
- [5] 罗永莲, 贾玉芳. 项目反应理论在题库建设中的应用研究[J]. 计算机应用与软件, 2015(1): 86-88.
- [6] 于海霞, 刘竞杰, 王家骥. 基于项目反应理论自适应考试系统的设计与应用[J]. 合肥学院学报(自然科学版), 2010(3): 44-48.
- [7] 丘威, 钟治初, 黄建妮, 等. 在线自适应测试系统的设计与实现[J]. 计算机应用研究, 2008, 25(1): 184-186.

作者简介:

单美静(1979-), 女, 博士, 副教授. 研究领域: 新型网络犯罪和计算机取证.

刘 琴(1975-), 女, 副教授. 研究领域: 数据挖掘.