

# 基于离散化误差补偿的试题难度题量控制方法

张 葵

(武汉科技大学计算机科学与技术学院 武汉 430081)

**摘 要** 自动组卷是高校实现考试规范化、科学化的重要手段。考生的平均成绩可以通过试卷的平均难度来控制。然而,如何确定各种难度的题量是关键技术。利用正态分布来确定各种难度的题量,是目前研究的一个方向。提出了对难度进行离散化正态分布后,利用对称积分求取各种难度的概率分布,并且采用比例误差补偿方法减少误差。同时,根据自动组卷的具体问题,确定了标准差取 0.2 更加合适,并且给出了在不同的总题量下各种难度题量的分配方案。结果表明,基于离散化误差补偿的试题难度题量控制方法能够较好地控制各种难度的题量,为自动组卷策略提供了有效的依据。

**关键词** 自动组卷,离散化,误差补偿,难度分配,正态分布

**中图法分类号** TP39 **文献标识码** A

## Discrete Error Complementary Strategy Based Control Method to Quantify Various Difficulty Levels of Test Questions

ZHANG Kui

(College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China)

**Abstract** It is extremely important to auto-generation examination papers for high school, with the purpose of making examination normative and scientific. The average scores of students can be controlled by average difficulty level of the examination paper. However, how to quantify various difficulty levels of test questions is a key step. Currently, a common strategy is to adopt normal distribution to specify those numbers. After discretizing the normal distribution function based on the difficulty factor, the probability distributions of various difficulty levels were obtained according to symmetrical integral and a proportion error complementary method was adopted to reduce errors. According to the analysis results in the auto-generation examination papers, the standard deviation of 0.2 is more suitable. In last, the computed numbers of various difficulty levels of test questions are listed in various total questions. The result shows that the control method that use discrete error complementary strategy can control the quantity of various difficulty levels of test questions well and enable the strategy of auto-generation examination papers to be effective.

**Keywords** Auto-generation examination papers, Discretization, Error complementary method, Difficulty distribution, Normal distribution

## 1 引言

目前,高校的试卷组织工作主要是人工组卷,其已越来越不能满足高校对试卷科学严谨、规范化的要求了。因此利用计算机来自动组卷在高校的教学工作中呼之欲出。

在自动组卷中,必须合理地设计试卷中试题的难度分布,使学生得分很高或低,都是失败的。在经典测量理论中,试题的难度可以通过试题的得分率或失分率来衡量,本文以得分率来衡量。学生的平均分可以通过控制整份试卷的平均难度来确定。也就是说,如果希望学生的平均分数是 60 分,那么试卷的平均难度系数可以定为 0.6;如果希望学生的平均分数是 75 分,那么平均难度定为 0.75。

目前自动组卷的算法得到了很多的研究<sup>[1-3]</sup>,但是在试卷中如何设计合理的难度分布,却没有得到足够的重视。因此,本文从这个角度来研究这个问题。

## 2 难度分布的研究现状

从教育测量学角度来说,考生的成绩应大致呈正态分布,这是设计合理的难度分配的理论基础<sup>[4]</sup>。因此各种难度题目的数量也应该是呈正态分布的。试题难度一般用难度系数来表示,值越大表示试题越难。现实中,难度系数不可能表示成连续变化的实数,需要将难度系数进行离散化<sup>[5]</sup>。文献<sup>[6]</sup>将考试成绩作为正态分布的横轴,并将其平均分为 4 个区间,分别对应 4 种难度级别:0~25 分对应难度 A“难”,25~50 对应难度 B“较难”,50~75 对应难度 C(较易),75~100 对应难度 D(易)。为了计算各个难度所对应的题量,分别在 4 个区间做积分,算出其所对应的概率,即为各个难度在总题量中所占的比例。当平均分(即期望值)为 50 分、方差为 10 时,算得的各个难度所占的比例为:难度 A 占 35%,难度 B 占 15%,难度 C 占 15%,难度 D 占 35%。但是这个结果显然与理论的

情况不符,如图 1 所示。从图 1 可以看出,0~25 所覆盖的面积显然远远低于 25~50 所覆盖的面积。

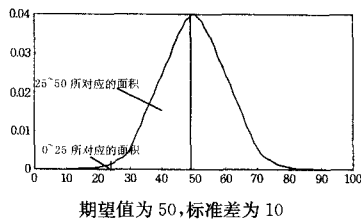


图 1 考试成绩正态分布概率密度图

文献[7]以试题难度作为正态分布的横轴,由于题目难度和题目数量均不能为任意实数,因此不能直接利用正态分布的概率计算公式来求各种难度所占的比例。所以该文献以离散化方式求取难度级别所占的比例,然后使用近似取整和修正算法求出试卷中各个难度所对应的题量,并以期望值为 0.5、方差为 0.3 计算出各种难度的题量。正态分布表达的是以难度为参数的概率密度,但是该文献将概率密度作为概率直接求解,将概率密度和概率混为一谈,因此其缺乏数学理论的支持。

基于上述算法存在的问题,本文提出了对以难度为参数的正态分布概率密度函数进行离散化,根据数学定义计算概率,然后进行误差补偿,以确定各种难度的题量的方法,并且以实验来比较、确定各种方法和参数。

### 3 基于误差补偿和正态分布的题量控制方法

若希望试题的难度呈正态分布,则难度  $x$  需要满足

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x=0.1, 0.2, \dots, 0.9 \quad (1)$$

式中,难度  $x$  由难到易分别取值 0.1~0.9,期望值  $\mu$  即为平均难度,标准差  $\sigma$  决定了分布的幅度。 $\mu$  越大,试卷难度越大; $\mu$  越小,试卷难度越小。 $\sigma$  越大,试卷难度越分散,抽中各种难度级别的可能性越均衡; $\sigma$  越小,试卷难度越集中,抽中离期望值近的难度级别的概率远远大于离期望值小的难度级别。标准差分别为 0.1、0.2、0.3 和 0.4 时的正态分布曲线如图 2 所示。从图 2 可看出,当  $\sigma=0.2$  或者 0.3 时,试题难度分布比较合适,因为当  $\sigma<0.2$  时,极难的试题和极简单的试题都不易被随机选中,而当  $\sigma\geq 0.4$  时,极难的试题和极简单的试题又太容易被随机选中。

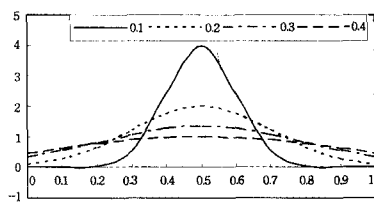


图 2 期望值为 0.5,标准差取 0.1~0.4 时的正态分布曲线

#### 3.1 难度概率的计算方法

以难度作为参数的正态分布函数,离散化后,不能直接利用式(1)计算  $f(x)$  的值来作为难度的概率。它们是不同的概念。由于难度的步长值为 0.1,因此根据计算概率的数学定义,可以在长度为 0.1 的区间上对正态分布进行积分,所得结果即为各种难度对应的概率。

积分区间长度为 0.1 时,计算难度  $x$  所对应的概率,可以有 3 种求法:

a)  $p(x) = \int_{x-0.1}^x f(x) dx$ , 称之为向前积分;

b)  $p(x) = \int_x^{x+0.1} f(x) dx$ , 称之为向后积分;

c)  $p(x) = \int_{x-0.05}^{x+0.05} f(x) dx$ , 称之为对称积分。

#### 3.2 基于误差补偿的各种难度级别所对应的比例

离散化过程的误差主要来自量化误差与截断误差,分析如下:

(1)形成序列的离散化过程产生截断误差。由于难度级别只取区间[0.1,0.9]的值,因此需要利用离散化数据窗对整个函数截取并抽样,得到近似的概率性序列,因此产生截断误差。

(2)离散化步长导致的量化误差。函数离散化后,先要求出各个难度所对应的概率,才能根据这个概率来分配各个难度的题量。由于离散化步长值为 0.1,将连续的空间离散化成抽样的值,因此积分不可避免地带来量化误差。

对随机变量离散化时,由于离散化区间长度有限,会出现“概率空间泄漏”,即离散化所得序列概率之和不等于 1。如果不进行处理,将导致后续序列运算时与实际结果间出现较大误差,因此序列离散化后,出现的误差需要进行补偿。文献[8]提出了两种补偿的办法,分别是平均补偿法和比例补偿法。

平均补偿法是将泄漏部分的概率平均分配至原序列中各非零元素上。假设补偿后的难度级别概率为  $p'(x)$ ,则

$$p'(x) = p(x) + \frac{1 - \sum_{i=0.1}^{0.9} p(i)}{9}, x=0.1, 0.2, \dots, 0.9$$

比例补偿法是将泄漏部分的概率,按照序列中各非零元素占序列各元素之和的比例加到各非零元素上,即

$$p'(x) = p(x) + \frac{p(x)}{\sum_{i=0.1}^{0.9} p(i)} \left(1 - \sum_{i=0.1}^{0.9} p(i)\right), x=0.1, \dots, 0.9$$

$$\text{误差补偿后, } \sum_{i=0.1}^{0.9} p'(x) = 1.$$

#### 3.3 近似取整和修正算法

假设总题量为  $n$ ,则各种难度所对应的题量为:

$$m_x = \lfloor n * p'(x) + 0.5 \rfloor (\lfloor \cdot \rfloor \text{表示取整})$$

近似取整后,由于存在舍入误差,必须加以修正。修正方法是:先求出误差,然后依次从平均难度开始向两侧难度修正题量,减少或者增加 1 个数量,直到与要求的总题量一致。

### 4 实验

文献[9]在考生分数呈正态分布,最高分 $\leq 98$ 的概率为 99.7%的假设下,提出了考分期望值的合理值域为[65,81]。基于这一理论,实验中将平均难度分别定为 0.65、0.7、0.75 和 0.8。

在提出的基于误差补偿和正态分布的题量控制方法中,如何选择合适的概率积分方法、如何进行误差补偿,以及如何确定标准差参数,是 3 个关键技术,决定了提出的方法是否能够更好地满足试卷的难度要求。

#### 4.1 概率积分方法的选择

实验方法:假设标准差为 0.2,不进行误差补偿,直接根据不同难度的概率求出难度  $x$  的概率所占的比例,然后使用近似取整算法求出不同难度所对应的题量  $m_x$ ,再计算出平均

难度,并与理论值做比较,求出相对误差。平均难度的计算方法是:

$$pjnd = \frac{\sum_{x=0.1}^{0.9} xm_x}{\sum_{x=0.1}^{0.9} m_x}$$

不同积分方法所对应的相对误差柱状图如图 3 所示。每张图中列出了总题量分别为 10、20、30、60 和 100 时的结果,柱子从左到右分别对应向前积分、向后积分和对称积分。从图中可看出,大部分情况下对称积分产生的误差比其他的积分方法产生的误差都小,因此本文选择对称积分方法来下面的实验。

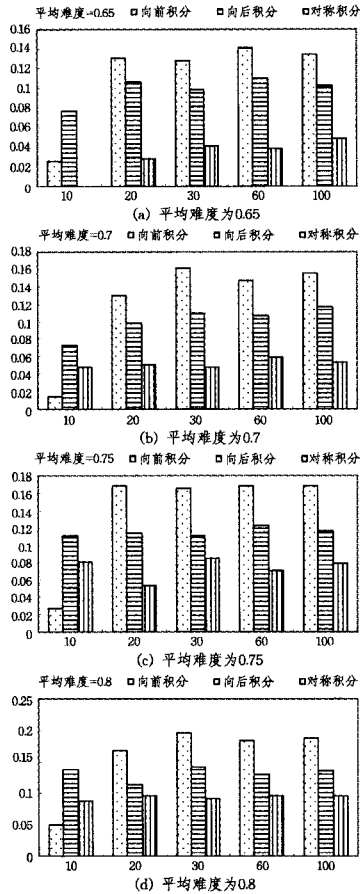


图 3 不同积分方法所对应的相对误差

#### 4.2 误差补偿方法的确定

误差补偿方法分为平均补偿和比例补偿。实验中,标准差为 0.2,概率用对称积分计算。不同平均难度的相对误差如图 4 所示。每张图中列出了总题量分别为 10、20、30、60 和 100 时的结果,柱子从左到右分别对应平均补偿和比例补偿。从图中可看出,比例补偿后的相对误差比平均补偿后的相对误差小。

#### 4.3 标准差参数的确定

由前所述,标准差取 0.2 或者 0.3 比较好,文献[6]中的标准差取为 0.3,实验后发现,标准差取 0.2 更好。实验结果如图 5 所示。实验中概率分布用对称积分计算,通过比例补偿方法补偿误差。平均难度分别取 0.65、0.7、0.75 和 0.8。每张图中列出了总题量分别为 10、20、30、60 和 100 时的结果,柱子从左到右对应标准差分别为 0.3 和 0.2。从图中可看出,标准差为 0.2 时,相对误差更小。

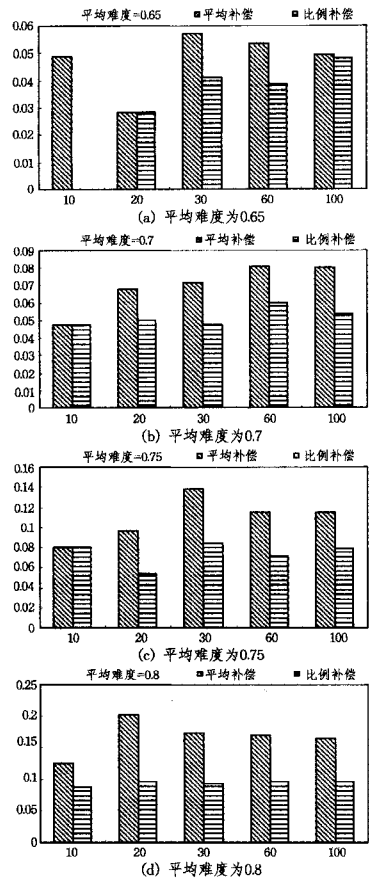


图 4 不同误差补偿方法所对应的相对误差

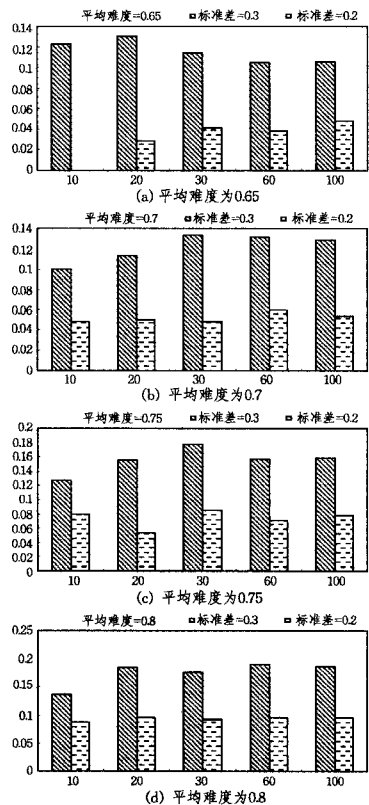


图 5 不同标准差  $\sigma$  所对应的相对误差

#### 4.4 不同题量下难度的分配方案

在上述各个结论下,采用近似取整和修正算法,题量分别为 20、30 和 60,针对不同的平均难度,计算出难度的题量分

配结果,如表 1 所列。

表 1 近似取整和修正算法对题量为 20,30,60 的难度分配结果

总题数	平均难度	难度								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
20	0.65	0	0	1	2	3	5	4	3	2
	0.7	0	0	1	1	3	4	4	4	3
	0.75	0	0	0	1	2	4	4	5	4
	0.8	0	0	0	1	2	3	5	4	5
30	0.65	0	1	1	3	5	6	6	5	3
	0.7	0	0	1	2	4	6	7	6	4
	0.75	0	0	1	2	3	5	7	7	5
	0.8	0	0	0	1	3	5	7	7	7
60	0.65	0	1	3	6	10	12	12	10	6
	0.7	0	1	2	4	8	12	13	12	8
	0.75	0	0	1	3	7	11	14	13	11
	0.8	0	0	1	2	5	9	14	15	14

**结束语** 利用正态分布来确定各种难度的题量是自动组卷问题中的关键技术,应该引起教育工作者的重视和研究。学生的成绩呈正态分布,可以通过控制平均难度来控制学生的平均成绩。在各种难度的题量控制研究中,主要有 3 个方面的研究内容:(1)难度不可能为任意实数,如何进行离散化,并以数学理论为指导计算各种难度的概率;(2)离散化后出现的误差如何补偿;(3)正态分布的标准化参数如何确定。以难度作为自变量的正态分布,离散化后,可以采用向前积分、向后积分和对称积分来求取各种难度的概率,其中对称积分所

产生的误差更小。可以采用平均补偿和比例补偿方法来减少误差,其中比例补偿的效果更好一些。通过多次实验,我们发现在自动组卷问题中,标准差  $\sigma$  取 0.2 会使计算结果更接近理论值。

## 参考文献

- [1] 张晓博. 基于 SOA 题库自动组卷系统的设计与开发[D]. 武汉:武汉理工大学,2011
- [2] 武仁杰. 蚁群算法在自动组卷建模中的应用研究[J]. 计算机仿真,2011(8):380-384
- [3] 刘毅. 人工智能在自动组卷建模中应用研究[J]. 计算机仿真,2011(8):385-388
- [4] 刘敏钰,李萍,杨丽军,等. 成绩分布正态性检验的研究与实现[J]. 新西部(下半月),2007(4):242-243
- [5] 王琴竹,孟朝霞. 通过试题难度控制试卷难度方法的探讨[J]. 电脑知识与技术,2009,5(3):761-763
- [6] 毛秉毅. 一种计算试卷中试题难度分布的有效方法[J]. 计算机工程,2002,28(6):280-281
- [7] 张殿龙. 正态分布函数在试题难度分配算法中的研究[J]. 哈尔滨理工大学学报,2009,4(13):13-15
- [8] 徐玮,康重庆,夏清,等. 序列运算离散化过程中的误差成因及补偿[J]. 清华大学学报:自然科学版,2009,49(1):17-20
- [9] 宣仲良. 各级难度试题分值比例的确定[J]. 淮阴师范学院学报:自然科学版,2004(3):24-25

(上接第 509 页)

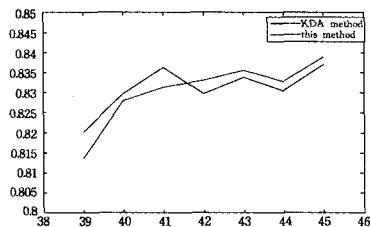


图 3 核函数  $k(x,y)=(x \cdot y+1)^2$  下在 XM2VTS 上的识别率

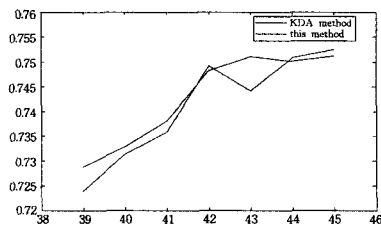


图 4 核函数  $k(x,y)=\exp(-\frac{\|x-y\|^2}{20})$  下在 XM2VTS 上的识别率

**结束语** 本文提出一种非参数核函数的判别分析方法(NPKDA),它通过在类间离散度矩阵中加入一个权值函数,使得类间离散度矩阵由类中心构造而成而非所有训练样本,其目的是为了地分析多个像素间的非线性相关性,从而有效地提取图像的局部特征并合理地利用各种不同类别的边界信息。该方法的实验结果表明,其识别率较原方法虽未有大的提高,但算法的稳定性有所加强,且可避免使用繁琐的矩阵奇异值分解理论,一定程度上简化了分类器的设计。但由于 KDA 仍然使用的是 Fisher 判别准则,因此在分类率方面,其并不是一种最优的方法。而判定一种核方法的优劣,关键因素在于所选取的核函数,因此核函数的选取是解决此类问题的关键。核选取对 KDA 的性能有较大影响。如何选择有关核函数,还需进一步的研究。

## 参考文献

- [1] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces; Recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):711-720
- [2] Mika S, Ratsch G, Weston R J, et al. Fisher discriminant analysis with kernels[C]//IEEE Neural Networks for Signal Processing Workshop. Wisconsin, USA, 1999
- [3] Li Zhi-feng, Tang Xiao-ou, et al. Nonparametric discriminant analysis for face recognition[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2009, 31(4):755-761
- [4] 郑宇杰,杨静宇,吴小俊,等. 一种新的核线性鉴别分析算法及其在人脸识别上的应用[J]. 计算机科学,2006,33(7):223-226
- [5] 田捷,杨鑫,等. 生物特征识别理论与应用[M]. 北京:清华大学出版社,2009
- [6] 徐东,刘政凯. 数据降维算法的研究及其应用[D]. 合肥:中国科学技术大学,2005
- [7] 王伟,张佑生,方芳. 人脸检测与识别技术综述[J]. 合肥工业大学学报:自然科学版,2006,2(29):158-162
- [8] Turk M, Pentland A. Face recognition using eigenfaces[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'91). Washington, 1991:586-591
- [9] Tefas A, Kotropoulos C, Pitas I. Using support vector machines to enhance the performance of elastic matching for frontal face authentication[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(7):735-745
- [10] Hua Yu, Jie Yang. A direct LDA algorithm for high-dimensional data with application to face recognition[J]. Pattern Recognition, 2001, 34(10):2067-2070
- [11] Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5):1299-1319