

在线考试防抄袭考生试卷生成模型^①

黄海明, 周琪云, 梁晓鹏, 张文字

(江西师范大学 计算机信息工程学院, 南昌 330022)

摘 要: 为了增大统一组织的在线考试中考生试卷间题目排列的差异性, 防止邻座考生相互抄袭, 提出了题目位置相似度的定义, 并根据位置相似度提出并实现了由同一份试卷, 生成多份题目排列顺序不同的考生试卷(变换卷)的模型. 通过与随机排列试卷题目方法考生试卷对比, 结果表明, 在按照题目类型和难度分组时, 使用该模型, 邻座考生间试卷题目排列的差异性明显大于随机方式.

关键词: 在线考试; 抄袭; 位置相似度; 变换卷; 差异性

Anti-Plagiarism Paper Generation Model for Online Examination

HUANG Hai-Ming, ZHOU Qi-Yun, LIANG Xiao-Peng, ZHANG Wen-Yun

(College of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China)

Abstract: To increase the diversity of question item arrangement between examination papers in an uniformly organized online examination, this paper proposes a definition of position similarity between examination papers, presents and implements a transformed paper generation model based on position similarity. Experiments show that this model could generate papers with larger diversity on question item sequence compared with the one generated by random arrangement when question items are grouped by item types and difficulty levels.

Key words: online examination; plagiarism; position similarity; transformed paper; diversity

在线考试^[1]在组卷、评判、统分等许多环节都减少了人工干预, 使考试变得更加灵活高效, 但在考试舞弊行为^[2,3]的预防和监控方面也比传统考试更难. 除了传统的舞弊行为外, 在线考试还衍生了使用网络传递答案等舞弊行为. 针对在线考试抄袭行为, 有些在线考试系统按照某种组卷策略^[4-6]为每位考生生成不同的试卷; 文献[7,8]提出了使用随机抽题的方式产生考试试卷; 文献[9]提出了通过监控USB端口来防止舞弊行为; 文献[10]提出了基于 Ajax 的监考系统; 文献[11]提出了通过监控网络中的数据包的方式检测舞弊行为; 文献[12]提出了基于 Silverlight 技术的防舞弊策略.

为每位考生生成不同的试卷的方法很难确保考试的绝对公平; 而使用网络或 USB 监控等方法能够监测考试中舞弊行为, 但并不能有效地预防抄袭行为发生.

本文提出并实现了一种变换卷生成模型, 让所有考生使用由同一份试卷(基准卷), 通过生成题目和选项排列顺序不同的变换卷来防止抄袭. 该模型的最大特点是能让邻座考生之间的试卷题目排列差异性最大.

1 变换卷生成模型

1.1 变换卷生成原则

本文中的基准卷为由人工出题或基于某种组卷策略生成的试卷, 变换卷是由基准卷按题目顺序重新排列生成的试卷. 使用变换卷的目的是确保在每位考生试卷难度完全相同的情况下, 增加考生之间试卷题目排列的差异性, 从而降低考生考试时相互抄袭的可能性, 以提高考试的公平、公正性. 变换卷的生成遵循以下原则:

① 不同考生试卷的题目排列顺序和选项顺序不同.

① 基金项目:江西省电子信息产业振兴和技术改造项目(2013260)

收稿时间:2014-09-20;收到修改稿时间:2014-11-03

② 邻座考生的试卷题目的排列顺序差异性最大, 即试卷题目的平均位置相似度最小。

③ 尽量使每个考生的试卷试题的排列顺序都不同。

④ 试卷照题目类型分组, 并且相同题型之题目按照难度等级从低到高排序。

1.2 位置相似度

根据对考试过程的观察与分析发现, 如果一个考生想要看身边考生试卷第 k 题的答案, 在看 k 题的过程中也可能会看到第 $k+1$ 题或者第 $k-1$ 题, 而看到其他位置的题目概率要小很多。因此认为, 如果同一道题目在两张试卷中题号相同, 或者题号差为 1, 则认为该题在上述两张试卷中的位置相似, 否则不相似。若把一份是试卷看成是试题按照一定顺序排列的数组, 则可以使用数组的特性来刻画试卷之间题目排列的相似度。

设 S_i, S_j 分别为 n 个不重复的整数 $a_1, a_2, \dots, a_k, \dots, a_n$ 按照不同排列顺序组成的数组, a_k 在数组 S_i 和 S_j 中的下标分别为 d_i 和 d_j , 即 $S_i[d_i] = S_j[d_j] = a_k$ 。定义 a_k 在数组 S_i 和 S_j 中位置相似度为: 若果位置相同, 相似度为 1; 若位置差为 1, 相似度为 0.5; 否则相似度为 0。用函数表示为:

$$h(a_k, S_i, S_j) = \begin{cases} 1, & |d_i - d_j| = 0 \\ 0.5, & |d_i - d_j| = 1 \\ 0, & |d_i - d_j| \geq 2 \end{cases} \quad (1)$$

其中 $n > 0, i \geq 0, j \geq 0$ 。数组 S_i 和 S_j 的元素平均位置相似度用函数表示为:

$$g(S_i, S_j) = \frac{1}{n} \sum_{k=1}^n h(S_i[k], S_j, S_j) \quad (2)$$

由于数组 S_i 和 S_j 中的元素相同, 根据函数(2)的定义可以知道 $g(S_i, S_j) = g(S_j, S_i)$, $g(S_i, S_j)$ 的取值范围为 $[0, 1]$, 当且仅当数组 S_i 和 S_j 中元素顺序完全相同时 $g(S_i, S_j) = 1$ 。

在文中规定, 由 n 道题目组成的基准卷 P_0 , 对应的数组为 $S_0 = [1, 2, \dots, n]$, 通过对数组 S_0 中的元素重新排列实现试卷题目的重新排列, 从而生成变换卷, 如果两份变换卷 P_i, P_j 对应的数组分别为 S_i 和 S_j , 那么这两份变换卷题目的平均位置相似度为 $g(S_i, S_j)$ 。

1.3 康托展开

康托展开是一个全排列到一个自然数的双射。其实是计算当前排列在所有由小到大全排列中的顺序。

康托展开公式为:

$$X = a[n] * (n-1)! + a[n-1] * (n-2)! + \dots + a[i] * (i-1)! + \dots + a[1] * 0!$$

其中, a_i 为整数, 并且 $0 \leq a[i] < i, 1 \leq i \leq n$ 。

因为康托展开是一个双射, 那么一定可以通过康托展开值求出原排列。根据康托展开值求原排列的过程称为康托展开的逆运算。

1.4 数组变换算法

根据变换卷生成原则, 提出了一种数组变换算法。该算法的功能是根据试卷题目数量和考生人数, 由一个全局整形二维数组 $kt[1, 2, \dots, 20][1, 2, \dots, m]$, 循环多次生成 m 个变换数组。

算法 1. 全局数组 kt 初始化算法

Input: m 考生人数最大值

Output: $kt[1, 2, \dots, 20][1, 2, \dots, m]$

LEN(arr)

return 数组 arr 中的元素个数

MIN(a, b)

return a, b 中值较小的一个

INVERSEKT(x)

$S \leftarrow x$ 通过康托展开的逆运算得到的排列中, 元素按顺序排列组成的数组。

return S

AVGSIMI(a, b)

$S1 = \text{INVERSEKT}(a)$

$S2 = \text{INVERSEKT}(b)$

return $g(S1, S2)$

INIKTARR(kt)

for $i \leftarrow 1$ to 20

arr[1] $\leftarrow 1$

$k \leftarrow \text{MIN}(m, i!)$

for $j \leftarrow 2$ to k

arr $\leftarrow k$ 不在数组 a 中, 并且使
AVGSIMI(arr[j-1], k) 的值最小。

for $j \leftarrow 1$ to m

kt[i] $\leftarrow$ arr[m mod LEN(arr)]

return kt

算法 2. 变换数组生成算法

Input: n 初始数组元素个数

m 生成变换数组个数

Output: $S[1, 2, \dots, m]$ 生成的变换数组

GREATEEXTARR(n, m)

```

 $S_o = [1, 2, \dots, n]$ 
if( $n \leq 20$ )
    for  $i \leftarrow 1$  to  $m$ 
         $S[i] \leftarrow \text{INVERSEKT}(kt[n][i])$ 
else
     $S[1] \leftarrow S_o$ 
    for  $i \leftarrow 2$  to  $m$ 
         $S[i] \leftarrow$  随机排列  $S_o$ , 直到  $g(S[i-1], S[i]) = 0$ 
return S

```

1.5 模型实现

因为一份试卷可能包含有多种题型, 而且各种题型的排列顺序是确定的, 比如第一部分是单选题, 第二部分是多选题, 最后是判断题。同时, 每个题目都有难度系数, 并且通常相同类型题目之间的排列顺序是按照从易到难的顺序排列, 该模型在实现时将难度分为简单、中等和难 3 个等级。

模型的实现步骤如下:

- ① 对基准卷的题目按照题目类型和难度等级进行分组;
- ② 对每个分组的试题生成与考试人数 m 相等的子变换卷数量: 假设基准卷的某个分组 P 中的题目按照题目编号从小到大排列为 $[a_1, a_2, \dots, a_n]$, 根据该分组中题目数量 n 和考试人数 m , 按照数组变换算法生成 m 变换数组, 每一个变换数组对应一份子变换卷。
- ③ 循环 m 次, 第 i 次循环中依次从每个分组中取第 i 份子变换卷, 组成一份变换卷。
- ④ 将生成的 m 份变换卷的题目顺序保存在数据库中;
- ⑤ 考试开始后, 登录系统, 获取与电脑 IP 地址对应的变换卷;
- ⑥ 将单选题和多选题选项顺序随机排列。
- ⑦ 返回试卷(变化卷)并显示给考生。

需要说明的是, 考场的计算机按照 IP 地址值从小到大进行排列, 并且相邻计算机 IP 地址值相差为 1。考试时, 将 IP 地址最后一段值对变换卷数量进行求模操作得到对应的变换卷编号。例如参加考试的人数为 50, 那么系统将生成 50 份变换卷。若果为考生的 IP 地址为 192.168.0.43, 那么该考生将会获得第 43 份变换卷。因为顺序相邻的变换卷对应试卷题目平均位置相似度最小, 所以相邻考生试卷题目排列差异性会最大。

2 实验验证

2.1 实验过程与结果

为了验证该模型生成的不同变换卷之间题目排列的差异性, 模拟实际考试中的情况进行了如下 2 个实验。

实验 1. 一份试卷题目总数量为 50 的试卷 A。题目类型和难度等级分布如表 1 所示。生成 6 变换卷。为了进行对比, 一组使用该模型生成, 另一组使用随机法对试题顺序随机排列生成。用函数(2)计算每份与前一一份试卷题目平均位置相似度, 结果如表 2 所示。

表 1 试卷 A 题目分布情况

	简单	中等	难
单选	10	12	8
多选	4	3	3
判断	4	4	2

表 2 相邻变换卷试题排列相似度对比($n=50$)

	1	2	3	4	5	6
该模型	-	0.06	0.09	0.06	0.09	0.06
随机法	-	0.39	0.29	0.38	0.33	0.40

实验 2. 一份试卷题目总数量为 100 试卷 B。题目类型和难度数量分布如表 3 所示。一组使用该模型生成, 另一组使用随机法对试题顺序随机排列生成。用函数(2)计算每份与前一一份试卷题目平均位置相似度, 结果如表 4 所示。

表 3 试卷 B 题目分布情况

	简单	中等	难
单选	20	24	16
多选	8	6	6
判断	8	8	4

表 4 相邻变换卷试题排列相似度对比($n=100$)

	1	2	3	4	5	6
该模型	-	0.00	0.01	0.00	0.01	0.00
随机法	-	0.22	0.12	0.16	0.13	0.12

2.2 实验分析

从上述两个实验结果可以看出, 在按照题目类型和难度等级分组的情况下, 当试卷题目数量为 50 时, 使用该模型生成的变换卷, 相邻两份试卷之间的题目排列相似度不到 0.1。若使用随机排列的方法, 相邻两份试卷的题目排列相似度大约在 0.35 左右; 当题目数量为 100 时, 使用随机变换的方法生成变换卷, 相邻两份试卷之间的下降到 0.15 左右, 而使用该模型, 相似度接近于 0。因此使用该模型生成的变换卷可以显著降低邻座考生之间的试卷题目的平均位置相似度,

即使邻座考生之间试卷题目排列差异性增大。

3 结语

本文提出了一种试卷题目位置相似度的定义和计算方法,根据位置相似度的定义提出并实现了一种变换卷生成模型,该模型可以根据题目的数量、题型、难度、参加考生的人数,由同一份试卷生成多份考生试卷。实验表明该模型生成的变换卷,邻座考生之间的位置相似度非常低,可以有效的预防在线考试中抄袭行为的发生。

参考文献

- 1 韩萌,王移芝,王嘉树.基于 Web 的通用在线考试系统的设计与实现.计算机工程与设计,2008,29(21):5608-5610.
- 2 黎敏.大学生考试舞弊现象分析及德行教育探讨.教育学术月刊,2009,7:52-53.
- 3 赵国年,张建军.大学生考试作弊成因及对策分析——以内蒙古农业大学为例.内蒙古农业大学学报(社会科学版),2012,14(4):98-101.
- 4 贺荣,陈爽.在线组卷策略的研究与设计.计算机工程与设计,2011,23(6):2183-2186.
- 5 周艳聪,刘艳柳.遗传模拟退火智能组卷策略研究.计算机工程与设计,2011,23(3):1066-1069.
- 6 齐万华,赵政文.基于蚁群算法思想的智能考试系统模型研究.计算机应用研究,2013,30(3):775-778.
- 7 李超燕,周建良.中小学信息技术上机考试系统的设计与实现.计算机应用与软件,2012,29(5):295-297.
- 8 李伟为.基于 ASP.NET2.0 的在线考试系统的设计与实现.计算机应用与软件,2011,28(9):163-165.
- 9 严强.在无纸化考试系统中实现 USB 端口监视.计算机应用与软件,2011,28(10):297-300.
- 10 范新龙,张华.基于 AJAX 技术的监考系统的设计与实现.计算机应用与软件,2011,28(9):188-190.
- 11 武伟,魏晓,魏仕民.基于监控服务器的在线考试防作弊方法的研究.计算机工程与设计,2007,28(8):1941-1943.
- 12 胡世清,程国雄.基于 Silverlight 防舞弊计算机网络考试系统的设计和实现.电化教育研究,2010,12:47-51.