

试卷模板的格元结构

谈云康

(四川铁道职业学院, 四川成都 611732)

摘要:为实现对相同知识点的不同试卷进行自动化编制,使用试卷模板智能生成结果试卷,采用格元结构(lattices and elements structure, LES)对试题的逻辑数据作存储和描述。通过定义和设计 LES 的类型、取值和算法,从而构造试卷模板,可以实现用软件灵活生成试卷的目的。

关键词:试卷模板;格;元;格元;LES

中图分类号:TP301.6

文献标志码:A

doi:10.16836/j.cnki.jcuit.2021.01.007

0 引言

利用微信平台将试卷发给学生在手机上解答,适时进行在线考试,经常需要重新对试题进行组织。命题组卷以人工为主,试卷知识点、考题类型等组卷因素需要人为把控。为实现自动化组卷,多数课题以传统的数据库组卷系统进行研究,建立动态试题数据^[1],或者用云计算技术构建智能化考试题库管理系统^[2]。

利用混合智能算法自动生成测试数据^[3],引入信息元素机制有效地保证搜索过程的多样性^[4],对组卷问题进行研究的主要方法有:基于遗传算法的优化数据处理技术;基于分段整数编码、多点交叉的改进遗传算法^[5];使用数据预处理算法模型去除数据中的错误^[6];按照题型和知识点联合分段编码^[7];基于概率论和自适应遗传算法的智能抽题算法^[8];不相邻试卷分配的补遗随机算法^[9];采用猎潜算法解决智能组卷问题^[10]等。这些方法基于试题库基础,其结构的建立可以使考试变得更加标准和客观化^[11],但存在组卷效率低下的弊端^[12]。

参与研究的对象是建立在试题库的基础上通过选题、组题和组卷算法而实现组卷不同,应用有序集理论^[13],将实体结构的元素表示为离散模型的集合^[14],采用的策略在于:1. 操作的对象不是题库而是一个模板;2. 组卷方法不是选题而是改题;3. 实现途径不是选择对象而是构建相似对象。

1 LES 的模型和作用

1.1 试卷模板的概念

定义1 设 P 是一份试卷,试题 $Q \in P$,当 Q 的一

个数据 d 改变后, Q 仍然成立,但 Q 的答案一定会改变,则称 d 为 P 的最小数据单元。

定义2 设相同知识点的试卷组成的集合为 S ,任意试卷 A, B ,有 $A \in S, B \in S$,且 $A \cap B = M$,则 M 在 S 中的补集 G 称为试卷集 S 的逻辑数据。

定义3 设相同知识点的试卷组成的集合为 S ,存在一个逻辑数据 G 不空。如果 $d \in G$ 对任意的试卷 $P \in S$,都有 d 是 P 的最小数据单元,则由 d 组成的数据结构称为试卷模板(test paper template)。

设 n 份相同知识点的试卷组成集合为 $S, S = \{A_i\}$ ($i=1, 2, \dots, n$),其中的每一个元素 A_i 是一份试卷,那么,任意两个 A_i ($i=1, 2, \dots, n$) 和 A_j ($j=1, 2, \dots, n$) 对应的不同部分只在数值上不同。对每一个 A_i ($i=1, 2, \dots, n$),把这些数据按序依次标为 N_k ($k=1, 2, \dots, m$)。

所有的 N_k ($k=1, 2, \dots, m$) 构成一个有序集 $L^s, L^s = \{E_r\}$ ($r=1, 2, \dots, m$)。由于 L^s 是一个有序集,所以 L^s 的取值可用一个数组 R_k 表示, $R_k = (E_1, E_2, \dots, E_m)$ 。

显然 A_i ($i=1, 2, \dots, n$) 和 A_j ($j=1, 2, \dots, n$) 有同样数量的试题数。假设每一道题有且仅有一个 N_k ($k=1, 2, \dots, m$),那么每一个 A_i ($i=1, 2, \dots, n$) 包含的试题数也是 m 。

对于 R_k 来说,每一个元素 E_r ($r=1, 2, \dots, k$) 的可取值及其个数与具体试题有关,任意两个元素 E_r ($r=1, 2, \dots, k$) 和 E_t ($t=1, 2, \dots, k$) 的取值个数并不一定相同。在实际问题中,不妨假定取值个数为相同的 w ,这时构成取值矩阵 $E = \{E_{ur}\}_{m \times w}$ 。

可见,矩阵 E 的每一个列矩阵对应一份试卷的 N_k ($k=1, 2, \dots, m$) 值;矩阵 E 的每一个行矩阵对应一个试题的 R_k 值 E_r ($r=1, 2, \dots, k$);矩阵 E 的任一元素 E_{ur} 一定存在于一个试题中。但是, L^s 中的任一元素 E_r ($r=1, 2, \dots, m$) 并不一定在矩阵 E 中,因为它的所有元

素按照上述法则组成一个非规范列的矩阵 $E^s = \{E_{uv}\}_{m \times w}$, 它的元素 $E_{uv} (u=1, 2, \dots, m; v=1, 2, \dots)$ 的列标 v 可能为无限。所以矩阵 E^s 是一个试卷模板, L^s 是 E^s 的一个数组描述。

1.2 LES 的概念

为组织一个试卷模板, 有必要对 L^s 界定一个数学模型, 称为“格元结构”(lattices and elements structure, LES)。

定义4 设 T 是一个试卷模板, $d \in T$, d 的一个数据特征 e 称为 d 的元(an Element)。

定义5 设 e 是一个元, 存储 e 的一个结构 L 称为格(A Lattice)。

定义6 设 e 是一个元, L 是一个格, e 及其格 L 统称为格元。

定义7 设 T 是一个试卷模板, $d \in T$, d 的格元按序组成的有序结构, 称为格元结构, 简称 LES。

1.3 LES 的结构

LES 的结构包括 3 部分: 标识符(identifier)、对象(object)以及属性(attribute), 如表 1 所示。

表1 LES 的结构

字段1	字段2	字段3
标识符	对象	属性

表 1 的每一部分是一个格, LES 包含 3 个格。每一个格中的对象是一个元, 所以标识符是 LES 的一个元, 称为标识符元(elements of the identifier)。

LES 的标识符元的标识对象是数据结构本身, 标识符元区别于模板文件中的其他元素, 在模板文件中具有唯一性, 无法和其他语言符号相同。LES 的属性值是整个结构的主体, 它指明 LES 的对象的类型。LES 的对象值(value of an object)是生成文件可取值的全集。

1.4 LES 的作用

从形式上看, LES 是其区别于生成文件的一个特征, 在模板文件中肯定有 LES 存在, 而在生成文件中一定没有 LES 存在。生成文件是由模板文件导出来的, 实际上是模板文件表达的一个子集。模板文件包含了所有的生成文件的类型。显然, 当模板文件的 LES 含有无限集的时候, 生成文件的份数也是一个无限集。

LES 不仅确定了生成文件的数量, 而且确定了生成文件的内容。首先, LES 应该定义生成文件的变量

值; 其次, LES 的属性值类型确定了生成文件的数量。

2 LES 的构建

2.1 标识符

标识格的元为标识符。标识符明确标示 LES 的格元(elements in the lattices, EL), 表示模板文件中要操作的数据对象。标识符是 LES 的一个元, 用“[[]]”表示, 中间没有空格。

标识格元只存在于模板文件中, 且不能存在于生成文件中, 否则, EL 的识别将不会唯一, 这会导致生成过程产生错误。原因是程序在处理时会将不是 EL 的字段进行修改, 使模板文件定义的内容被误操作, 产生非预期的结果。因此, 生成文件的内容是不能含有和标识符相同的字符的, 如果有, 应该用其他字符替换, 这点在制作模板文件时应该特别注意。

2.2 对象

对象格是 LES 的主体部分, 格中的元称为对象元(elements of an object)。对象元被标识符包含, 如[[3]], 其中的对象元是 3。

对象格列举或描述对象元, 每一个对象元作为一个选择项提交到生成文件中。在[[3]]中, 3 是一个对象元; 而在[[3, 2]]中, 有两个对象元分别为 3, 2。

2.3 属性

属性格是对象格的一个描述, 指明对象格的元属性。属性格的作用是为算法提供指标, 算法根据这个指标为对象元赋值。

属性元(elements of an attribute), 是指操作对象的取值类型。分两类: 变量型和枚举型。

2.3.1 变量型

变量型, 其值可用一数集表示。如[[R]], 对象元可取实数集中的任一值。R 表示实数集, I 表示整数集。

2.3.2 枚举型

枚举型, 其值可取列表中的任一个。如[[3, work, 4.5]], 对象元可取“3”或“work”或“4.5”中的任意一个值, 每个值用逗号“,”隔开。单独的一个值为枚举型, 如[[3]]。

3 对象元

3.1 对象元算法

(1) 提取属性值的类型;

(2) 将变量型赋值;
(3) 取得枚举型值。
设枚举型存入数组中,如表 2 所示。

表 2 枚举型数组数据				
字段 1	...	字段 $i+1$...	字段 $n+1$
A_0	...	A_i	...	A_n

对于任一 $A_j(j=0,1,\cdots,n)$, A_i 是集合 $\{A_j\}$ 中第 $i+1$ 个元素。取整数集 $\{0,1,\cdots,n\}$ 中的一随机数 i ,得到 A_i ,作为枚举型值。

3.2 对象元算法的实现

这里只就与对象元提取相关的算法作一简要介绍。

先用当前时间作为函数的初始化种子数
 $\sin(\text{Random})=\text{Timer}()$
再用此值构造一个负初始化值,作为系统函数 $\text{Rnd}()$ 的输入
 $\sin(\text{Random})=\text{Rnd}(-\sin(\text{Random})-3)$
对 $\text{Rnd}()$ 的初始化完成后,用空值函数取得目标值

$\text{intI}=\text{Int}((\text{int}(\text{Count})\times\text{Rnd}())+1)$
上面的做法看起来 intI 与 $\sin(\text{Random})$ 并不相关,为什么要这样做呢?

首先,当 $\sin(\text{Random})=0$ 或 $\sin(\text{Random})>0$ 时, $\text{Rnd}(\sin(\text{Random}))$ 的每个第 1 次调用的返回值相同,多次调用的返回值组成一个数列,其返回数列也相同,如下所示:

0.7055475,0.533424,0.5795186,...;0.7055475,0.533424,0.5795186,...;...

其次,当 $\sin(\text{Random})<0$ 时,对于每一个固定的 SeedValue 种子数,多次调用函数 $\text{Rnd}(\text{SeedValue})$ 取得的值是一个常数列;对于不同的种子数 $\sin(\text{Random})$,取得不同的常数列,如下所示:

0.7133257,0.7133257,0.7133257,...;0.7055475,0.533424,0.5795186,...;...

所以需要用一个随时变化的初始种子数 $\sin(\text{Random})$,其值为负,取得初值;然后,再用空值调用 $\text{Rnd}()$,取得其余各个所需值。

4 生成试卷的结构分析

由于主观题的已知条件构造灵活,可以不受限制,而客观题的已知条件要和答案项匹配,所以重点分析客观题。

为便于统计,以单项选择题为例:每个选择题 4 个常规选项,1 个容错选项,构造 10 个题的模板,使用通过本课题的理论编制的软件 AutoMakerEQT 生成 4 套题。以题序为横轴、每题的正确选项为纵轴作出散点图,如图 1 所示。

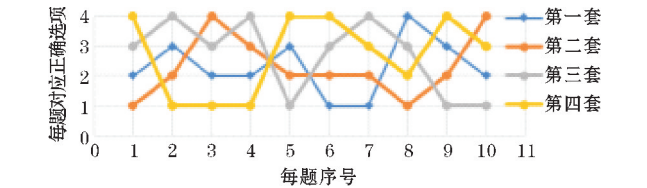


图 1 单选题答案的散点图

由图 1 可知,生成的结果试题符合组卷要求。

5 结束语

对于每个 LES,随机取得的枚举型值 $A_i(i=0,1,\cdots,n)$ 可能有重复。从 $A_j(j=0,1,\cdots,n)$ 中取得的 A_i 对应 i 值属于整数集 $L:\{0,1,\cdots,n\}$,其循环周期 $T=n+1$ 。设试卷的生成次数为 N ,对于每个 LES,可以用以下方式处理重复:当 $N\leq T$ 时,每次取得 i 值后,同时将其从数集 L 中删除,下一次从新的 L 中取得 i 值;当 $N>T$ 时,初始化数集 L 以后循环操作。显然,当 $N\leq T$ 时, i 值不会重复。

对于每个 LES,随机取得的变量型值 $A_i(i=0,1,\cdots)$ 也可能有重复。从 $A_j(j=0,1,\cdots)$ 中取得的 A_i 对应 i 值属于整数集 $L:\{0,1,\cdots\}$ 。每次取得 i 值后,同时将其从数集 L 中删除,下一次从新的 L 中直接取得 i 值,由于没有循环周期,不受生成次数 N 的限制, i 值不会重复。

构造试卷的过程由软件实现不再详述。

经过试验,通过模板建构生成文件,只要合理使用 LES,就能得到所需要的结果试题。

参考文献:

[1] 张恒强,吴琼,杨博超,等.大学无机化学题库建设—简答题题库的组建[J].山东化工,2019,48(21):224-227.
[2] 余战秋,文晶晶.基于 Hadoop 云计算的在线考试题库管理系统设计与研发[J].鄂州大学学报,2019,26(6):98-100.
[3] 夏春艳,张岩,万里,等.基于否定选择遗传算法的路径覆盖测试数据生成[J].电子学报,2019,47(12):2630-2638.

- [4] 于笏韵,刘传才.一种面向路径的测试数据自动生成的组合方法[J].计算机与数字工程,2019,47(8):1951-1955.
- [5] 潘刚,杨清平,蒲国林,等.遗传算法在智能组卷系统中的应用研究[J].云南民族大学学报(自然科学版),2016(6):579-584.
- [6] 常征,吕勇.基于正则表达式的海量数据清洗系统[J].计算机应用,2019,39(10):2942-2947.
- [7] 席卫文,张春辉,王飞,等.一种基于改进遗传算法的医学题库自动组卷设计与实现[J].中国医学物理学杂志,2016,33(8):861-864.
- [8] 石中盘,韩卫.基于概率论和自适应遗传算法的智能抽题算法[J].计算机工程,2002,28(1):141-143.
- [9] 孟祥娟,王俊峰,曹锦梅.利用遗传算法实现试题库自动组卷问题[J].计算机系统应用,2010,19(1):180-184.
- [10] 何福男,陈树越,史国栋.猎潜算法及其在智能组卷中的应用[J].信息与控制工程,2011,26(3):86-90.
- [11] 王玉刚,董友亮,陈晔,等.《航空军械装置》课程动态试题数据库的建设及研究[J].数据挖掘,2018,8(1):17-22.
- [12] 韦忠庆,袁煜,蔡军.基于SaaS模式的分布式在线考试系统设计[J].现代电子技术,2019,42(12):48-51,56.
- [13] Jukka Kohonen. Generating Modular Lattices of up to 30 Elements[J]. Order: A Journal on the Theory of Ordered Sets and its Applications, 2019,36(3):423-435.
- [14] Mijo Nikolić, Emir Karavelić, Adnan Ibrahimbegovic. Lattice Element Models and Their Peculiarities[J]. Archives of Computational Methods in Engineering. 2018,25(3):753-784.

Lattices and Elements Structure of Test Paper Template

TAN Yunkang

(Sichuan Railway College, Chengdu 611732, China)

Abstract: To automatically generate different test paper with the same knowledge points, the test paper template is used to generate the result test paper intelligently, the logic data of the test questions are stored and described by using the lattices and elements structure(lattices and elements structure, LES). By defining and designing the type, value and algorithm of LES to construct the test paper template, the purpose of using software to generate the test paper flexibly can be realized.

Keywords: test paper template; lattice; element; lattices and elements; LES