

# 基于多层过滤和动态概率模型的试题抽取算法研究

傅 勉<sup>a,b</sup>

(安徽新华学院, a.商学院; b.大学生素质教育研究中心, 合肥, 230088)

**摘要:**针对研究生考试涉及课程门类众多, 内容要求灵活多变, 并且存在数门课程合并出卷的特点, 提出一种将多层过滤模型和动态概率模型相结合的试卷抽取方法, 首先通过多层过滤模型层层过滤出符合出题要求的试题, 然后使用动态概率模型对所有试题按照一定概率进行动态调整。结果表明试题分布合理, 完全能够满足研究生考试试题抽取的需求。

**关键词:** 多层过滤; 动态概率模型; 试题; 抽取算法

**中图分类号:** TP301.6    **文献标志码:** A    **文章编号:** 1673-1891(2019)02-0059-04

## Research on Graduate Entrance Examination Questions Extraction Algorithms Based on Multilayer Filtering and Dynamic Probability Model

FU Mian<sup>a,b</sup>

(a. Business School, b. Research Center for Quality Education; Anhui Xinhua University, Hefei, Anhui 230088)

**Abstract:** With regard to the characteristic diversity of subjects, updated flexible contents, and unified examination papers for multiple courses in Unified National Graduate Entrance Examination, a random examination questions extraction algorithm based on multi-layer filtering model and dynamic probability model is proposed. First, the multi-layer filtering model is used to filter out the examination questions that meet the requirements for the examination, and then the dynamic probability model is used to adjust all the examination questions dynamically according to a certain probability. The results show that a reasonable distribution of questions has been achieved, which can well meets the requirements for the examination questions extraction in the National Graduate Entrance Examination.

**Keywords:** multilayer filtering; dynamic probability model; examination question; extraction algorithm

## 0 引言

当前, 随着计算机智能信息技术的实施及普及, 研究生单考中试题抽取的智能性、灵活性、可靠性成为高等学院教育信息化发展的一项关键课题。然而, 从目前研究生单考试卷抽取的情况来看, 试卷抽取随意, 没有完全考虑到研究生出题的特点和灵活性, 对于试卷抽取质量也没有合理的评价<sup>[1-3]</sup>, 因此, 需要针对研究生试卷出题的特点, 设计专门的试题抽取算法。论文针对安徽新华学院研究生单考课程考试的特点, 首先通过多层过滤模型层层过滤出符合出题要求的试题, 然后使用动态概率模型对所有试题按照一定概率进行动态调整, 从而实现对研究生试题抽取的科学性和精确性。

## 1 试卷生成特点分析

### 1.1 研究生考试个性化需求分析

研究生考试因为涉及门类较多, 而且多门课程复合出卷, 存在不同于其他类型考试的个性化需求, 具体表现在以下几个方面:

(1) 当前研究生招生不仅学科多, 而且专业门类设置也比较多, 各个专业需要考试的科目也比较多, 导致每年招生考试中出题环节十分复杂, 教师工作量不小。

(2) 研究生考试包括统考科目和单独考试科目两大类, 这两大类对于试卷出题的要求不相同, 增加了出卷老师的出卷难度。

(3) 某些专业课之间需要复合出题, 即出一张试卷, 每一门课程的出题数量、所占分数要求都不

同,这需要出题老师之间的沟通和合作。

(4)在研究生试卷中,基于优秀学生选拔的需求,会有一部分拔高题,这些题目主要考察学生对于课程中一些知识点的综合理解能力和灵活运用能力,这类题目的卷面分数要求不得多于总分的20%。

(5)每一年的研究生试卷会根据情况发生变化,主要表现在考核知识点侧重性以及考核的具体范围和内容上。

(6)考虑到研究生考试的特殊性和保密需求,要求试卷的题目具有客观性和保密性。

### 1.2 需要解决的问题

鉴于这种情况,在研究生试卷自动生成系统中,如果可以构建一个比较科学的研究生试题库,构建试卷生成系统自动抽取试题,再辅以某种手工调整策略,就可以达到自动生成完全满足需求的试卷,使得研究生考试管理走上高效运转的轨道,将对试卷的正规化管理起到十分关键的作用。因此,建立研究生试卷自动抽取系统可以解决下面几个问题:

(1)研究生试卷的题目难度等级、题量多少教师难以把握,使得手工出试卷的质量普遍偏低。

(2)手工出卷中教师对于知识点的全面涉及及拔高题的设置容易出现重复或者缺失现象,导致测试效果不准确。

(3)由于教师自己出题可能需求经历比较长的时间,容易引入不必要的泄题可能,造成试卷的失密后果。

(4)由于是手工出卷,教师工作量十分大,容易出现这样或那样的错误。同时,可能出现某种题目多年出现在试卷中的现象。

## 2 组卷过程分析

为建立试题库,并实现试题的抽取、组卷,需要解决数据安全、流文件格式转换、数据存储、文档生成、抽取算法设计等问题,其中最为关键的问题就是试题的抽取算法,因为系统不但要求能对单门课程进行试卷的生成,而且还对多门课程进行复合生成试卷,对于算法设计来说一般要考虑到难度调节、分值均衡、知识点的筛选及重点知识考核、题型分配等问题。试卷的组卷过程如图1所示。

(1)首先根据出题人输入的考试课程,从试题库中筛选这门课的全部试题,将其存储在临时表里。

(2)再根据出题的题目类型,从临时表中筛选

出考试要求的类型题目,放在一个临时表中。

(3)继续从中筛选出符合出题人意图的试题,放入临时表中。

(4)重复步骤二和步骤三,一直到生成出题人满意的试卷为止。

(5)最后,系统提供对生成试卷的评价功能,可以根据评价结果对试卷进行局部调整,直到满意为止。

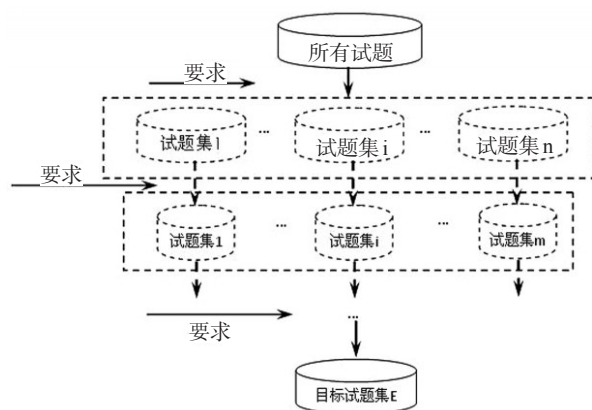


图1 组卷过程图

## 3 基于多层过滤模型的试题抽取

### 3.1 多层过滤模型算法思想

所谓多层过滤模型,就是按照出题人的出题课程、出题类型、知识点、难度系数要求,把某门课程的试题通过一层一层不断过滤,逐步得到满意试卷的过程。

多层过滤模型如图2所示,基于多层概率模型的抽取过程如下<sup>[4-5]</sup>:

(1)第一步,根据出题人输入的课程名称,从试题库中抽取这门课的全部试题,放入中间表 first\_table 中;

(2)第二步,根据出题人的考试试题类型要求,分别从中间表 first\_table 中抽取出各个类型的试题,放入中间表 middle-1、middle-2、...、middle-s 中;

(3)第三步,根据出题人设定的考核知识点,分别对中间表 middle-1、middle-2、...、middle-s 进行再一次筛选,抽取出符合要求的试题,放入中间表 m-1、m-2、...、m-s 中;

(4)第四步,根据出题人设定的难度系数,分别对中间表 m-1、m-2、...、m-n 进行再一次筛选,抽取出符合要求的试题,放入中间表 n-1、n-2、...、n-s 中;

(5)第五步,把中间表 n-1、n-2、...、n-s 的所有试题统一集中存储到 extraction\_table 表中;

(6)第六步,启动试卷评价功能,对系统生成的

试卷进行综合性评估,如果试卷满足设定的要求,则将相应试题写入 Word 文档;否则,可以进行人工干预,调整试卷。

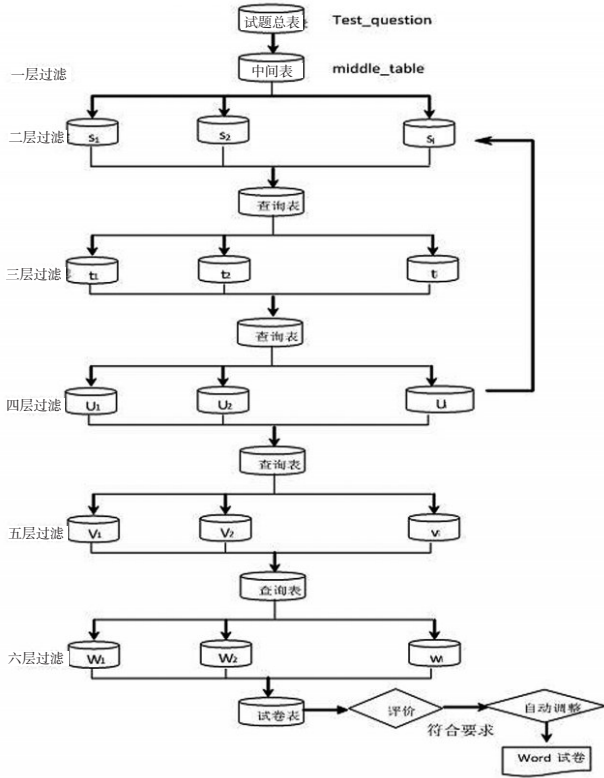


图2 多层过滤模型流程图

### 3.2 多层过滤模型算法描述

试卷抽取生成的过程算法如下<sup>[1]</sup>:

```
{stress_point:array[1..50,1..2] of integer;//知识点数组
```

```
point:array[1:50] of integer;
```

```
h:integer;//试卷平均难度
```

```
//六个层次的过滤
```

```
test_question->middle_table->s1,...,s2,s1;
```

```
s1,...,s2,s1->t1,...,t2,t1->u1,...,u2,u1;
```

```
u1,...,u2,u1->v1,...,v2,v1->w1,...,w2,w1;
```

```
while (stress_point[i]<>0)
```

```
{s1,...,s2,s1->t1,...,t2,t1;}
```

```
//针对重点知识点,抽取
```

```
while(point[i]<>0)
```

```
{s1,...,s2,s1->t1,...,t2,t1;}
```

```
//对剩下的各知识点,抽取
```

```
while (h满足要求) or (两次h相同)
```

```
{s1,...,s2,s1->t1,...,t2,t1;}
```

```
评价h;
```

```
if 满足要求 then
```

```
W1,...,w2,w1->exam_table->word 文档;
```

```
else
```

```
人工调整->word 文档;
```

```
}
```

## 4 动态概率模型

在前述的多层过滤模型中,需要经过多次的层层筛选,为了防止有些试题在历年考试中被使用的频率太大,保障考试的公正性和低重复性,需要采用某种方法设定题目的使用概率,这正是动态概率模型研究的关键所在<sup>[6]</sup>。

### (1) 试题被抽取概率的描述

我们假设结果集合为某门课程的被抽取的题目集合,设为 $\{q_1, q_2, \dots, q_n\}$ ,这个集合的抽取概率分别是 $\{p_1, p_2, \dots, p_n\}$ 。

①第一次抽取试题时,这个结果集合中每一个题目的抽取概率是一样的,即 $p_1 = p_2 = \dots = p_n$ ,这时我们可以按照系统产生的随机数来进行选取;②如果这个结果集合中的某些题目 $q_i$ 被抽取过多次,那么,其对应的抽取概率 $p_i$ 就需要及时变化。

### (2) 试题被抽取概率的动态调整

首先假定如下情况:某个考试题目被抽取过一次之后,如果再一次被抽取,那么我们人工增加该试题的抽取次数,并随着抽取次数的增加,降低该试题的抽取概率;如果该题目总的被抽取数目达到原先设定的 $m$ 次后,这时需要重新恢复该题目的原始概率。这么做的优点是既可以动态改变某个试题的抽取概率,防止某些试题反复出现,又不会一段时间内无法抽取到该试题。

除了采用定期对题库进行有计划的更新策略外,还采用扩大结果集这个策略保证抽取试题的可解性,具体方法是:已知试题集合中各个试题的已经

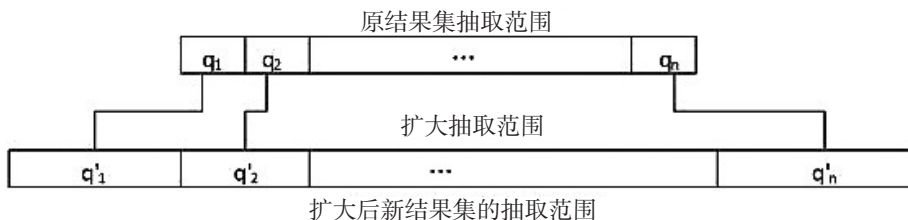


图3 扩大查找范围示意图

抽中次数,可以按照抽中次数的若干倍来扩大试题集范围,具体抽取次数与扩大倍数关系如表1所示:

表1 抽取次数与扩大倍数关系表

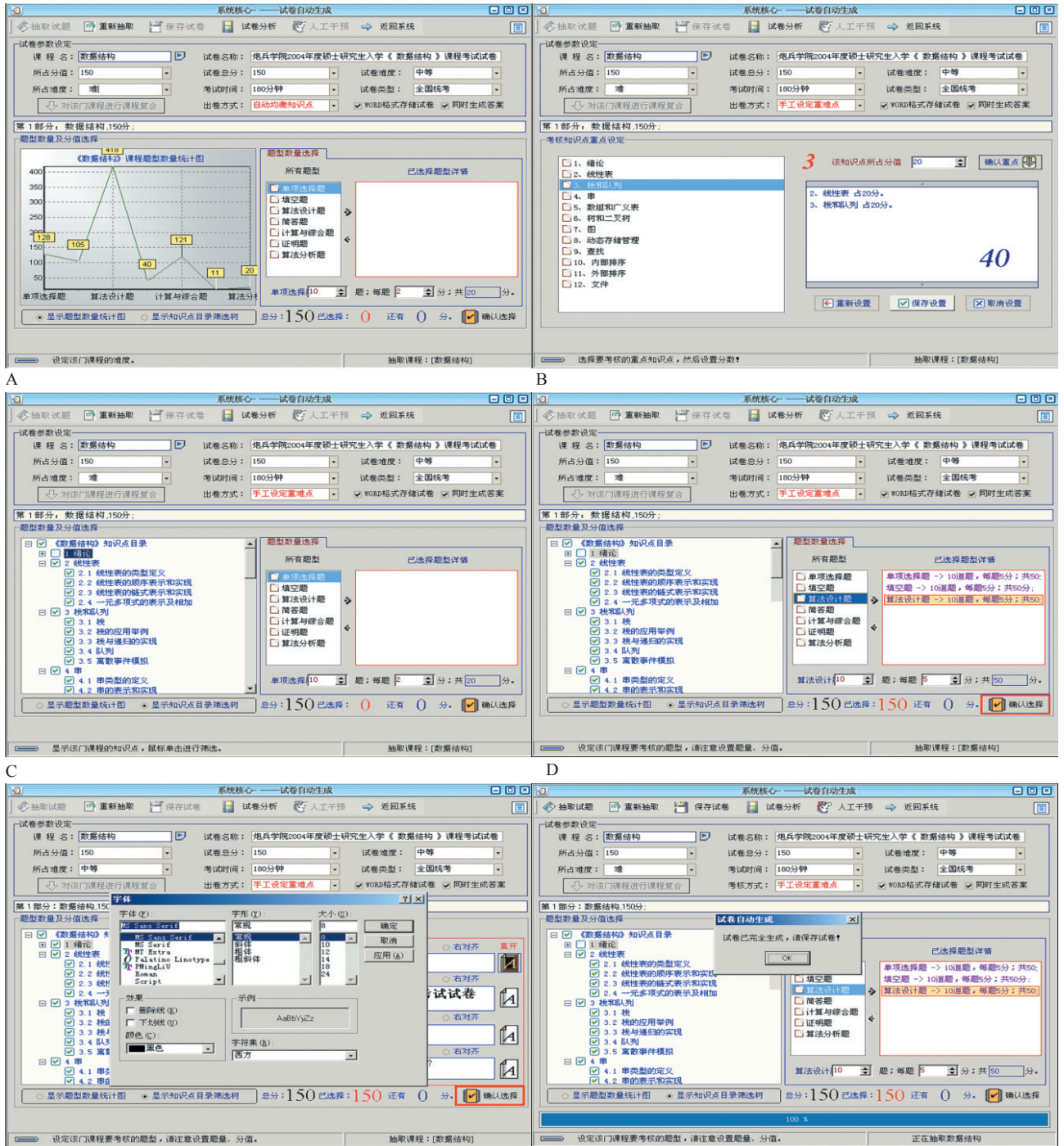
抽取次数	0	1	2	...	m-1
扩大倍数	nm	nm-1	nm-2	...	n1

可以根据具体情况明确m和n这两个参数值,在新的试题集中随机来抽取,如果抽取出来的结

果出现在某试题的扩展范围之内,那么该试题即被抽取,如图3所示。

### 5 试卷生成验证

以计算机课程“数据结构”为案例,利用系统生成该课程试卷的过程按图4中A~F的标号依次进行:



A 课程选择分值、难度设定、课程复合;B 重难点知识点的设定;C 屏蔽知识点操作;D 试题类型与分值的设置;E 试卷格式的设置;F 试题抽取及保存

图4 系统生成该课程试卷的过程

(下转第124页)

- [4] 武用鑫.AutoCAD机械制图实训教程[M].北京:北京邮电大学出版社,2016:33-39、57-61.
- [5] 林党养.AutoCAD2010机械绘图[M].2版.北京:人民邮电出版社,2014:177-185.
- [6] 高红英,刘军旭,袁惊滔.CAD与机械制图“融合式”教学改革探讨[J].成都工业学院学报,2015,18(4):82-85.
- [7] 吕波.工程制图[M].北京:北京邮电大学出版社,2016:89-154.
- [8] 王技德,王燕.AutoCAD机械制图教程[M].3版.大连:大连理工大学出版社,2018:173-275.
- [9] 王晨曦.机械制图[M].北京:北京邮电大学出版社,2017:20-22.
- [10] 张凤林,刘战涛.机械制图[M].哈尔滨:哈尔滨工程大学出版社,2016:110-116.

(责任编辑:曲继鹏)

(上接第62页)

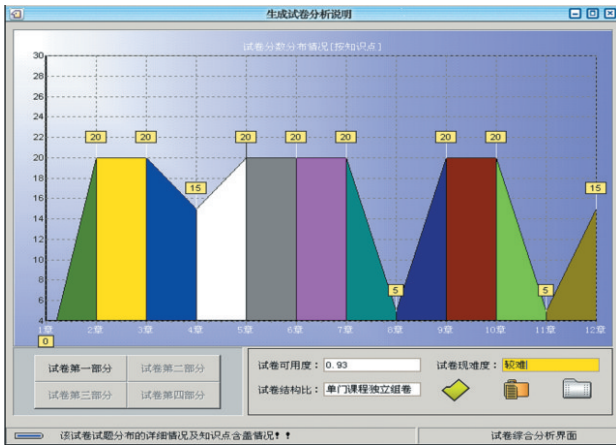


图5 试卷分析结果

对生成的试卷知识点和题型分布进行分析,按

照试题抽取算法所生成的试卷知识点分布及题型分布如图5所示。

可以看出,生成的试卷知识点分布中各个章节的题量分布比较均匀,相关性也适中。

## 6 结语

本文针对研究生考试不同于其他考试的独特性,根据出题人的出题需求,首先采用多层过滤模型对试题库进行层层筛选,实现对试题类型、知识点、考试重点的选取,然后通过动态概率模型对前期筛选出的试卷进行动态调整。该系统已经在多个高校的研究生考试中得到实践应用,结果显示,该系统能够满足研究生考试的出题需求。

### 参考文献:

- [1] 王凤蕊,王文宏,董金新.基于自适应果蝇优化算法的平行试卷自动生成[J].计算机工程与设计,2015,36(10):2807-2812.
- [2] 徐守萍.试题库系统及组卷策略[J].科学技术与工程,2007(14):3631-3634.
- [3] 孙荪,皇甫正贤.基于数据库的大型管理信息系统优化设计[J].计算机应用研究,2001(6):88-90.
- [4] 赵英杰,赵菲.计算机考试系统的开发应用[J].管理信息系统,2000(S1):32-35.
- [5] 赵英杰.试卷自动生成系统的开发与应用[J].管理信息系统,2000(S1):36-37.
- [6] 季玉茹,王德忠.题库与试卷生成系统设计[J].吉林化工学院学报,2008,25(6):97-99.

(责任编辑:曲继鹏)