

基于题库自适应在线考试系统的分析与设计

杨昱昺¹, 王红艳²

(1. 宁波大红鹰职业技术学院计算机与信息工程系, 宁波 315175; 2. 河北工程技术高等专科学校电气系, 沧州 061001)

摘要: 针对目前在线考试系统设计组卷方面存在的问题(如试题的难度系数不能随着被测试人群的变化而自动调整 and 同题型同知识点的试题重复出现等), 提出了创建一种以伯努利大数定理和机器学习思想等理论为基础的基于题库自适应在线考试系统。并通过对题库管理、出卷管理和考试管理3部分的介绍, 系统地阐述了基于题库自适应在线考试系统的设计与实现方法。

关键词: 题库; 自适应; 在线考试; 难度系数

中图分类号: TP 391

文献标识码: A

文章编号: 1671-4431(2008)04-0157-04

Based on the Question Bank Self-adapted Online Exam System Analysis and Design

YANG Yu-bing¹, WANG Hong-yan²

(1. Department of Computer Information and Engineering, Institute of Ningbo Dahongying Professional Technology, Ningbo 315175, China; 2. Electric Department, Engineering and Technical College of Hebei, Cangzhou 061001, China)

Abstract: Currently online exam system design, dealing with the key issues aspects of getting the test paper(such as the coefficient of difficulty of the questions can not change with the acceptability of the test groups change and the knowledge questions with repeated questions on the same test paper, etc.), proposed to create an online exam system based on the question bank by the theory of Bernoulli's theorem and machine learning etc. And by introducing three parts of the system of examination management, getting test paper management and test management, based on the data bank on Adaptive Online Examination System Design and Implementation. Systemic expound the method of design and realization of the online exam system based on the self-adapted question bank.

Key words: question bank; self-adapted; online exam; difficulty coefficient

随着科技进步和 Internet 的应用普及, 学校或教育认证机构为了提高教学质量, 严把考试关。实行考教分离和推行标准化考试制度是常用的手段, 而这二者的基础是需要一个比较完善的在线考试系统。该系统允许教师, 根据考试需要自动从题库抽取试题生成试卷, 并将试卷印发给学生进行标准化考试, 或直接在机房进行网络在线考试^[1-2]。

一个考试系统, 一般由自动和手动 2 种组卷方式。自动组卷方式是系统根据教师提供的信息(题量、难度、章节、考试时间等)进行自动组卷。对于进行网络在线考试方式, 试卷可采用一人一卷考试时自动随机组成, 也可采用多人一卷。手动组卷方式是在自动组卷的基础上进行人为调整。所以, 自动随机组卷的考试模

收稿日期: 2008-02-01.

基金项目: 宁波市教育局青年基金课题.

作者简介: 杨昱昺(1969-), 男, 副教授. E-mail: yybljf@hotmail.com

式是考试系统的基本模式。而这种模式的质量主要依赖于试题库试题的难度系数的准确程度,另外,避免同知识点的问题型试题重复对试卷的组成也有较大的影响^[3-5]。

目前,虽然已经有了不同种类的在线考试系统,但这些考试系统在一定程度上还存在这样或那样的问题。从组卷方面来看,主要归纳为如下几个方面:1)虽然试题的难度系数是由有经验的教师预先设定,但靠经验指定必定有误差,由于受主观和客观的影响,不可避免带来估计的偏差。2)难度系数不能随着考试群体接受情况的变化而改变,考试结果难以准确把握。3)问题型同知识点的试题重复(以下简称试题重复)出现情况比较严重。一般题库在构建时,对某个知识点,往往会有多个类型相同,内容稍作变化的试题相对应,但在组卷时,且不希望试题重复出现。

针对上述问题,根据伯努利大数定律^[6]、机器学习思想^[2]和人工辅助性调整,提出了基于题库自适应在线考试系统的分析与设计方法。

1 难度系数自适应学习

对于考试而言,假设考试者是相互独立的,而且是独立做出答案。设试题只有一个考核点,采用答对得分答错不得分的阅卷模式;又设某试题累计得分次数为 N_s ,该题被选中次数为 N ,则该题到目前为止的得分率为

$$q_k = N_s / N \quad (1)$$

根据伯努利大数定律^[6],当考试次数 N 趋于无穷时, q_k 将以概率 1 接近得分概率 q 。

一般情况下,很多试题都有多于 1 个的考核点,一道试题的得分概率是这道试题所有考核点得分概率的平均概率,即

$$q = \frac{1}{M} \sum_{i=1}^M q_i^s \quad (2)$$

式中, M 为该题的考核点数。

需要指出的是,一道试题的多个考核点有可能是线性相关的,但为了简化和计算需要,假设其是线性无关的,即独立的。

题库中的难度系数 d 可由得分概率 q 得到,即

$$d = 1 - q \quad (3)$$

难度系数自适应的处理方式如图 1 所示。

试题难度系数的初值由专家确定,以后在线测试模式下,每当学生考试结束后,将结果提交给系统,系统对学生提交的考试结果进行自学习,重新计算难度系数,并修改对应题库中的难度系数值,以适应当前学生的实际解题能力。随着考生人数和考试次数的增加,难度系数逐步逼近实际情况。

为了实现难度系数自适应,在题库表中必须加入如下字段:

Field1: 试题编号

Field2: 试题的题目

Field3: 试题的难度系数

Field4: 试题的考点数目

Field5: 本题累计被命中次数 N

Field6: 考核点 1 累计得分次数

Field7: 考核点 2 累计得分次数

.....

Field $M+5$: 考核点 M 累计得分次数(可设 M 为 5,因为一般一道试题的考点不会超过 5 个,若有特殊情况,可增加考核点字段)

2 避免试题重复自我调整

假设知识点 A 在题库中问题型试题有 X 题,而题库中问题型的试题有 N 题,那么当在同一份试卷中同

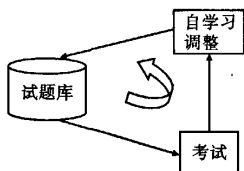


图1 系统自学习调整

题型同知识点的试题出现2题以上,则认为试题重复,这样系统自动出卷所发生的试题重复概率可用下式得到

$$q = \frac{X}{N} \sum_{i=2}^X \frac{i}{X} = \frac{(X+2)(X-1)}{2N} \quad (4)$$

若没有进行处理,则系统每次自动出卷时知识点A的试题重复概率不变,为了降低重复概率,可在系统自动出卷后,由教师在试卷中指出重复的试题,并提交,系统根据提交情况对重复试题按知识点分建立关联组,系统再次自动出卷时,将采取同一组内的关联试题最多只允许在试卷中出现一题,这样公式(4)将转换成如下形式

$$\left\{ q \mid q \leq \frac{(X+2-i)(X-1-i)}{2N}, i \in \text{自然数} \right\} \quad (5)$$

式中, i 是每次发生知识点A的试题重复,系统调整后改变的变量,初值为0,每发生一次, i 值加1。这样随着 i 的递增,知识点A在同一试卷中出现的试题重复概率逐渐减小,直至为0,也即对知识点A在题库中同题型试题的关联字段已全部做上关联标记。

避免试题重复自我调整的处理方式如图2所示。

关联字段的初值为空,当系统自动出卷后,教师发现有同题型同知识点的试题,系统允许教师指出并提交,系统实行自我调整,将教师指出的试题关联字段打上关联标志,当系统再次重组试卷时,只允许在试卷中留下其中一题,其余用字符串匹配法屏蔽,并从题库剩余试题中抽取。这样,随着系统的不断被使用,题库中所有同题型同知识点的试题被逐渐设定为关联,试题重复出现的出错概率也变得越来越大,最后将被全部消除。

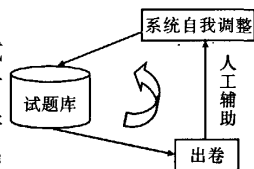


图2 人工辅助系统自我调整

关联标志,可采用如下方法:如试题 x_1 、 x_2 的题号(是唯一的)为 $xz000001$ 、 $xz000045$,则当系统发现两者并联后,在自我调整时会将 x_1 、 x_2 试题的关联字段值都更新为: $xz000001|xz000045$ 。

为了避免试题重复,在题库表中还应加入如下字段:

FieldM+6:与此题关联的试题。

3 基于题库的自适应在线考试系统的构成

系统主要包括题库管理、出卷管理和考试管理3部分。这3部分既相互独立,又相互联系。

1) 题库管理模块 该模块主要实现题库的建立和维护工作。主要有3大功能:(1)题库的建立。一般一门学科对应一个题库,首先对某一学科进行试题的录入,录入的试题应由经验的专家设定相应的初始值(难度系数、试题的考点数目)。(2)题库的维护。包括基本的管理功能(添加、删除和修改)及查询功能。(3)系统的管理。包括日常的管理工作(如用户的登录身份验证、数据存贮管理及事务处理等)。

2) 出卷管理模块 该模块主要完成在线组卷、实现题库避免试题重复自我调整 and 打印试卷。主要有3大功能:(1)在线组卷。教师可根据需要由系统自动组卷,或先由系统自动出卷,再通过教师根据系统所出的卷进行调整,并将调整结果提交给系统,如此反复,直到满意为止。(2)实现题库避免试题重复自我调整。系统自动出卷后,根据教师提交的反馈信息(难度系数的手工调整或指出同题型同知识点的重复试题),对题库进行调整,以适应应试者的实际情况。(3)打印试卷。对于实行非在线考试的情况,系统将打印出试卷,以供使用。在线组卷流程如图3所示。系统自动出卷子模块流程如图4所示。

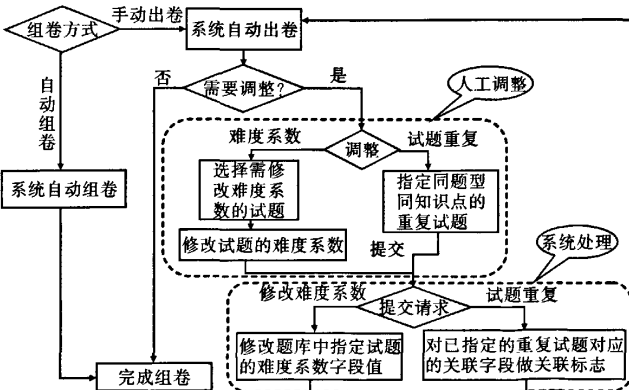


图3 在线出卷流程图

系统自动出卷子模块流程如图4所示。

3) 考试管理模块 该模块主要完成自动出题、呈题及实现题库难度系数自适应。主要有3大功能:(1)

自动出题。为了满足学生自我学习的需要,在该模块中允许学生自己出卷,自己考试。(2)呈题。在统一试卷考试中,学生获取已经由教师出好的试卷进行考试,考完后将结果提交给系统。(3)实现题库难度系数自适应。系统对学生提交的考试结果进行判定,对回答正确的试题,将题库中相应试题的有关字段值进行更新,更新的字段有考核点累计得分次数和试题的难度系数,并将成绩加到成绩累加器中,否则直接进入后面答题的判断,最后将成绩记录到学生成绩数据库中,供学生自己或老师查看。

在此求得的难度系数是一个 $[0,1]$ 之间的连续实数,但在实际处理时,往往将试题的难度级别设为5级:难、较难、适中、较容易、容易,将难度系数量化如表1所示。

表1 难度系数与难度级别的转换

难度级别	难	较难	适中	较容易	容易
难度系数	1	0.8	0.6	0.4	0.2

系统是采用阈值法进行转换,根据公式(2),求得得分概率,再通过公式(3),求得难度系数,再通过阈值法,将难度系数转换成5级。

题库难度系数自适应模块流程如图5所示。

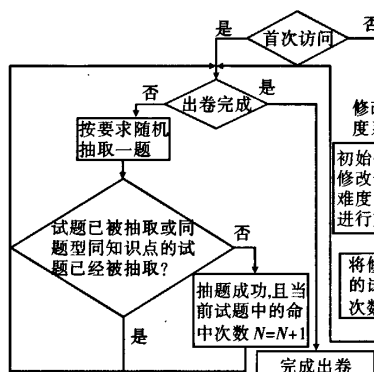


图4 系统自动出卷子模块流程图

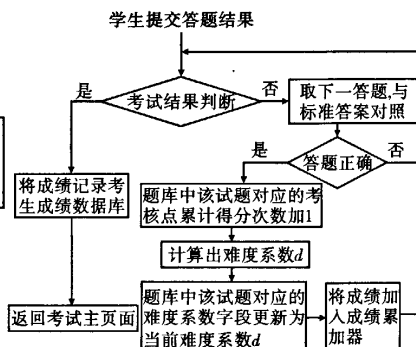


图5 题库难度系数自适应模块流程图

4 结 语

以目前在线考试系统出卷时经常出现的问题为入口,以题库难度系数自适应和避免试题重复自我调整为中心,从理论到实际,系统地阐述了基于题库自适应在线考试系统的分析与设计。该方案曾在宁波大红鹰学院在线测试系统改进版中使用,对非计算机专业大学生计算机等级测试和计算机专业C语言二级水平理论部分进行测试,达到了比较理想的预期效果。

参考文献

- [1] 成汝震,刘宏忠.基于Web自适应考试系统[J].计算机应用,2002,22(7):122-124.
- [2] Tom Mitchell. 机器学习[M].北京:机械工业出版社,2003.
- [3] 王晓荣,王 萌.基于自适应的智能组卷方案研究[J].荆州师范学院学报:自然科学版,2002,25(15):33-35.
- [4] 林雪明.试题难度系数确定数学模型的建立与实现[J].杭州应用工程技术学院学报,2001,13(1):47-51.
- [5] 李一波,张溶溶.试题得分概率和答题时间概率分布自适应学习整定[J].计算机工程与应用,2005,41(17):215-217.
- [6] 金炳陶.概率论与数理统计[M].北京:高等教育出版社,2000.