

基于 Web 的自适应性智能试题库的设计

Design of Web - Based Exam - question with Self - study and Adaptive Adjusting

刘丽平 (中国科学院研究生院 100049)
(北京市电子信息学校 100011)
王文杰 (中国科学院研究生院 100049)

摘要:随着网络技术和智能信息处理技术的发展,基于网络的智能教学系统(ITS)的研究逐渐成为热点问题。针对高等教育自学考试者知识掌握参差不齐及面对海量知识点往往束手无策的特点,本文设计了一个基于 WEB 的自适应题库系统。该系统可动态分析学习者的学习状态,将试题的难度系数动态自适应调整,再运用信息推送技术,有针对性地向学习者提供个性化的学习资料和信息。本系统在美国英语语言学院电子信息教学部自学考试辅导软件中使用、验证并达到了预期效果。

关键词:大数定律 难度系数 信息推送技术

1 引言

高等教育自学考试以其突破年龄限制和时限限制而日益成为学习型社会中继续教育、终身教育的重要形式,也成为不可或缺的获取国家文凭的方式,越来越多的普通院校开始开设自考部。随着网络技术和智能信息处理技术的发展,基于网络的智能教学系统(Intelligent Tutoring System, ITS)的研究也已经成为一个重要课题。目前有些 ITS 也设置了试题库,一些自考辅导部门也开发了智能计算机辅助教学软件 ICAI,但总存在这样或那样的不足,例如:

- (1) 试题库的智能性、自适应性不够。
- (2) 试题库试题的难度系数是由有经验的教师指定的,一经指定就不再变化。且不考虑靠经验指定有偏差,难度系数也不能适应考试群体可接受情况的变化。
- (3) 面对海量的知识点,学习者很难快速找到和定位自己所需要的学习资源。
- (4) 大多数试题库只给出参考答案,没有给出学生应巩固的知识点或学习策略的建议,也不能根据学习者自身学习状态自动组卷。

针对以上问题,为了适应二十一世纪个性化、多元化教育和服务性教育的需要,本文运用了人工智能的科学理论和方法,根据大数定律中的贝努利定理(Ber-

noulli theorem) 和机器学习思想,利用信息推送技术,提出了一种智能性、自适应性较强的高等教育自学考试题库的设计模型。

2 智能题库的设计目标

(1) 动态跟踪学习者的测试过程,根据暴露出的问题进行机器学习,自动调整试题库试题的难度系数,这样会使自动组卷的依据更科学。对于自考辅导机构的辅导教师也能根据不同群体的特征,调整教学重、难点,从而做到有的放矢。

(2) 根据考核结果,系统会把出错的知识点提取出来,运用信息推送技术,自动推送出相关的学习资料和测试资料,确保学习者能更好地掌握和巩固暴露出的学习瑕疵,有效解决了学习者学习资源的快速查找与定位,从而提高了学习者的学习质量和学习效率,更能适应个性化的学习要求。

3 基于 Web 的自适应性智能试题库的功能模型

本试题库的功能模型见图 1,各模块的功能介绍如下:

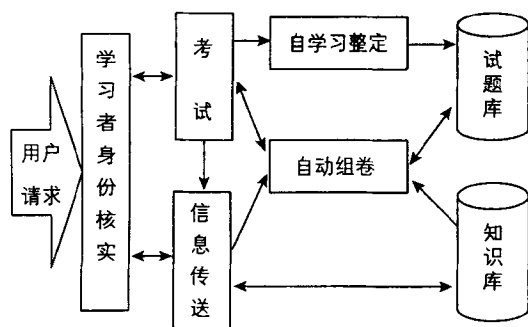


图 1 功能模型图

(1) 用户身份核实模块。负责接收用户名和密码,并核实学习者身份的合法性。

(2) 考试模块。负责接收学习者的考试请求,如请学习者选择考试科目、考试范围、题型、题量、时间等,并将该学习者的请求传送给自动组卷模块。在收到自动组卷模块发送来的试卷以后,负责维护用户的考试。用户提交试卷以后,该模块负责判卷,并将考试结果反馈给信息传送模块和自学习整定模块。

(3) 自学习整定模块。负责接收考试模块反映出的每道题的考核结果,据此调整试题库中相应试题的属性,并定期对试题库试题的难度系数进行整定。

(4) 自动组卷模块。负责接收考试模块传送来的用户组卷要求。组卷时,本模块首先顺序搜索各条规则,根据相应的规则索引符合条件的所有试题,再根据试题难度系数按比例随机生成试卷。另外,本模块还可根据信息传送模块传送来的学习者测试过程中暴露出来的知识点,结合用户提出的组卷要求生成再测试试卷,以利于学习者进一步的巩固和提高。

(5) 信息传送模块。负责将考试结果推送给用户并生成用户日志。本模块可根据日志访问知识库,将与失分点相关的知识从知识库中提取出来自动推送给用户,并将重测要求提交给自动组卷模块。利用基于网络的数据挖掘技术动态分析日志,本模块可向用户推出在线成绩簿、进步幅度图、学习进度表及需深入学习的知识点、相关的辅助复习资料等。

(6) 试题库模块。用于存储试题。试题库中的试题按不同题型分别存放在数据库中。试题按考核科目、知识点、难度系数的不同分类管理,这样可以很容易生成一份知识点分布均匀的试卷。

知识库模块:用于存储学科领域知识点。

4 关键技术的实现

4.1 难度系数自适应整定方案

为了实现试题难度系数的自动调整,要在试题库的试题中加入如下属性:

表 1 需要加入的属性及属性说明

属性名称	属性说明
难度系数级数	记录试题难度系数级别,如 10 级或 100 级。
绝对难度系数	由有经验教师的指定,一经指定就不再发生变化,以保持试题库的原始状态。
相对难度系数	是自动组卷的依据,初始值等于绝对难度系数,每隔一段时间(如一周)运行“自学习整定”算法,学习得到新的难度系数,从而使试题难度系数可随每届学生知识掌握的不同特征而变化。
命中次数	记录本题累计被命中的次数。
失分次数	记录本题累计失分的次数。

4.1.1 理论依据——贝努利大数定律^[1]

设 m 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 的概率: $p = p(A)$, 则对于任意正数 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} \left\{ \left| \frac{m}{n} - p \right| < \varepsilon \right\} = 1$$

贝努利定理表明:当 n 充分大时“频率 $\frac{m}{n}$ 与概率 p 的绝对偏差小于任意给定的正数 ε ”。就是说当 n 很大时,事件发生的频率与概率有较大偏差的可能性很小。由此推断原理,在实际应用中,当实验次数很大时,便可以用事件发生的频率来代替事件的概率。

4.1.2 难度系数自适应整定算法^[2]

由分析可知,难度系数正比于失分概率,所以可以通过分析失分概率或失分率得到难度系数。设采用答对得分答错失分的阅卷模式, P_i 为考核点的失分率, N 为累计答错次数, N 为该题被累计命中次数,则 $P_i = \frac{N_i}{N}$ 。随着考试次数的增加,根据贝努利定理, P_i 将依概率 1 收敛于失分率 p_i 。若一道题有多个考核点,则可将该题的失分率定义为该题所有考核点失分率的平均值。设 m 为考核点数, $P_{i,j}$ 为该题第 j 个考核点的失分率, P_i 为该题的失分率,则 $P_i = \frac{1}{m} \sum_{j=1}^m P_{i,j}$ 。

本系统中难度系数是有量级的,则难度系数与答错概率的关系依难度系数级数或者成绩级数量化一致。设 Δd_0 为量化难度系数,则 $\Delta d_0 = \Delta [P_0]$ 。

4.2 信息推送技术^{[3][4]}

随着互联网信息源以爆炸性的速度持续增长,人们在信息检索方面遇到了很多困惑。为了满足用户在网络上及时查找有用信息的需要,信息推送技术应运而生,并成为第三代浏览器的核心技术。推送技术正在以其提供服务的主动性、返回信息的新颖性、及时性,以及在个性化信息服务方面的优势而日益得到人们的重视,并成为新一轮信息资源利用、占有、控制的争夺热点。

4.2.1 推送技术概念

推送技术,就是根据用户的需求,有目的地按时将用户感兴趣的信息主动发送到用户的计算机中,即实现“信息找用户”。用户可在想看的时候再查看它,甚至可以离线浏览。这不仅改变了因特网上信息流动的方向,还通过信息流量的减少而降低因特网的负载,对用户也没有技术上的要求。

4.2.2 网上信息推送技术的四种形式

(1) 频道式推送。频道式网络推送技术是目前网上普遍采用的一种推送方式,它将某些网页定义为浏览器中的频道,用户可以像选择电视频道那样去选择收看感兴趣的、通过网络播送的信息。Microsoft、Netscape 等都有自己的频道定义格式。例如,Microsoft 公司提出的频道定义格式是为站点信息内容建立“目标索引文件”,以便于个人化定制信息的推送;Netscape 提出的则是基于“元内容”的网播方式。

(2) 邮件式推送。用电子邮件方式主动将有关信息发布给列表中的用户。

(3) 网页式推送。在特定网页(如某企业、某机构或某个人的网页)上将信息提供给感兴趣的用户。

(4) 专用式推送。通过机密的点对点通信方式,将指定的信息发送给专门的用户。Microsoft 和 Netscape 公司的 Web 浏览器均能支持上述各种信息推送方式。

4.2.3 信息推送在自适应智能试题库中的应用

网上信息推送技术的四种形式都可应用在本系统中,具体表现在以下几个方面:

(1) 系统将试题库使用说明、学习建议、相关复习

资料、高等教育自学考试优秀站点等多种信息以频道式推送方式推送给学习者。

(2) 利用网络信息挖掘中的 Web 日志挖掘技术,系统自动分析 Web 服务器的日志文件,对学习者访问 Web 服务器方留下的访问记录进行挖掘,从中得出用户的访问模式,进行个性化分析处理,如制作在线成绩簿、进步幅度图、学习进度表等,再以频道式推送方式或邮件式推送方式推送给用户,以便于用户了解自己的学习状态,并起到激励作用。

(3) 在每次测验完成后,系统会将学习者与失分点相关的知识点从知识库中提取出来,以网页式推送方式推送给用户,再根据学习者暴露出的知识点自动组卷,以网页式推送方式或频道式推送方式自动推送给用户进行再测试。另外,系统还可把历届考试真题每一份每一题都建立与知识点的联系,再将其涉及到这些知识点的考题以四种形式中任何一种形式推送出来。这些形式的信息推送,可以确保学习者能更好地掌握和巩固考核阶段暴露出来的学习瑕疵。

(4) 系统会动态分析学习者的学习日志,并为学习者推送出给出建设性意见和建议,如:需深入学习的知识点、相关的辅助复习资料等。

(5) 系统会将历届国家考试真题以邮件式推送方式推送给学习者,使学习者进一步明确课程的考核方式。

5 结束语

本系统在美国英语语言学院电子信息教学部自学考试辅导软件中被实现,经 3 年近 5000 人次累计使用,验证了其可行性,证明了其有效性,下一步的目标是利用信息融合技术,使试题的难度系数更科学的自适应不同的考试群体。

参考文献

- 1 盛能骤、谢式千等,概率论与数理统计[M],高等教育出版社,1997。
- 2 李一波、张森悦,试题库试题难度系数自适应学习整定[J],计算机工程,2005,31(12):181-182。
- 3 陆谊,信息推送技术在网络教学中的应用[J],计算机应用与软件,2005,22(5):65-67。
- 4 吴剑霞,信息推送技术的原理及在图书馆的应用[J],科技情报开发与经济,2004,14(3):19-21。