

基于知识的计算机辅助命题 及分析系统外壳的系统设计*

林复华 陈光中 洪潮兴 汪国强

(华南理工大学应用数学系, 广州 510641)

摘要 本文首先简述 GITB shell 的结构与功能, 继而着重提出以“指标体系”为引导的知识获取策略及成卷知识的层次结构和表示, 然后介绍了试卷设计的推理机制, 建立了面向试卷模式的试卷设计任务分解、分层随机启发式和冲突求解等求解策略。

关键词 计算机辅助命题, 知识获取, 专家系统, 系统外壳。

建立计算机命题系统属于 CAI 的范畴, 涉及到教育科学与计算机科学及其紧密结合。在国外, 虽有一些对标准化考试建立的试题库计算机系统, 例如著名的美国 TOEFL、GRE、GMAT 考试计算机系统以及新加坡的化学标准化考试计算机系统, 但是技术细节是保密的。近年来, 我国高校不少学科建立了试题库系统^[1-3]。这些系统多数属于管理型试题库系统, 其成卷质量参差不齐。为提高此类系统水平, 减少重复劳动, 在国家教委高教司倡导下, 从 1992 年 12 月起, 我们开始研制“全国高等院校课程通用智能成卷系统开发工具”(简称 GITB shell), 它是基于知识的计算机辅助命题及分析系统外壳。

由于命题是人类教学专家运用所掌握的知识 and 经验完成的复杂的一个从整体到局部、从模糊到精确、自顶向下、逐步完成的智能设计活动, 因此, 成卷专家系统是基于知识的设计型 (design-typed) 专家系统。系统能吸取、集中课程专家(群)的具有权威性的经验, 能分析、识别不同类型用户的模糊的或不完整的命题要求, 具有完备、灵活的推理机制。因此, 我们建立通用的知识表示格式、推理模型和运行环境, 它能反映考试学原理、不同课程的特点, 形成完整、一致的成卷知识体系。对任意一门课程, 只要在该课程专家(群)的合作下, 输入该课程的命题经验知识和结构分布合理的试题, 就可制作出该课程的计算机辅助命题系统。

本文首先简述 GITB shell 的结构与功能, 继而结合一些实例, 着重提出以“指标体系”为引导的知识获取策略及成卷知识的层次结构和表示, 然后介绍了试卷设计的推理机制, 建立了面向试卷模式的试卷设计任务分解、分层随机启发式和冲突求解等求解策略和实现技术。

* 本文 1994-06-14 收到, 1994-09-23 定稿

作者林复华, 1962 年生, 讲师, 现为香港科技大学博士研究生, 主要研究领域为工业工程学, 数学模型。陈光中, 1936 年生, 教授, 主要研究领域为 AI。洪潮兴, 1939 年生, 副教授, 主要研究领域为 CAI, 成卷理论。汪国强, 1940 年生, 教授, 主要研究领域为基础数学, CAI。

本文通讯联系人: 林复华, 广州 510641, 华南理工大学应用数学系

1 GITB shell 的结构与功能

GITB shell 以 Windows 3.1 作为开发平台和运行环境,采用 Windows 统一友好的图形界面(如弹出菜单、对话框、列表框、选中框等)及其对不同设备类型的适应能力,采用中文之星 Cstar1.2 以上汉字系统.系统主要分为知识获取与维护(系统生成)、成卷推理机制、测试及系统维护 4 个子系统.系统结构如图 1.

各子系统基本功能:

(1) 知识获取与维护

这部分功能是由于成卷系统制作中心,我们建立了以“指标体系”为引导(indexes-directed)的知识获取与维护策略及知识的层次结构和框架、“IF...THEN...”规则、判定表、过程的各种表示方法,具体为:

(a) 建立课程指标体系;用 Foxpro 2.5 for Windows 设计用户界面.

(b) 专家直接知识(包括:指标库、模式库、规则库、试题正文/解答库)的增加、删除、修改、查询、统计、压缩/解压等;试题正文解答采用 Word 2.0 for Windows 的格式(.DOC)用 Borland C++3.1 for Windows 开发编辑/打印模块;压缩/解压模块是基于目前压缩率最高的 ARJ 软件(Shareware)开发而成.

(c) 产生专家间接知识(包括选题指标确定次序,题库结构特征矩阵,分层随机启发式规则,误差界限);由系统与课程专家交互产生.

(d) 知识的一致性、完备性检验.完备性检验有:题库分布的基本要求检验;模式库、规则库的完备性检验.一致性的检验有:试卷模式与题库结构的相容性检验;指标、正文、解答三者间的对应检验;规则间的条件冲突检验.

(2) 成卷推理机制

可实现 4 种成卷方式:(a)智能全自动成卷;(b)带特殊要求的自动成卷;(c)输入部分或全部成卷模式的半自动成卷;(d)逐题输入指标的课程专家成卷.可交互地调整试卷(增题、删题、换题、调分),成卷结束后,产生试卷内涵覆盖表、试卷双向细目表、成卷情况分析表.

(3) 测试评价子系统.主要由测试分析模型库和考试反馈信息数据库组成.功能有

(a) 根据考试反馈信息,对试题、试卷作定量的统计分析,提供实测资料,用以完善试题指标体系和成卷知识;当把考试结果(抽样或全体)输入后,以报告表格形式打印输出成绩分布、成绩平均数、中位数、众数、成绩分布的正偏态或负偏态情况、成绩分布差异数、试卷难度、信度、效度、试题难度、区分度等.

(b) 对考试结果进行评判并分析教学效果.

(4) 系统维护主要功能

(a) 对“题库制作中心”,提供系统所使用的各类数据库、知识库提供显示、打印、增删、修改、转储等功能.

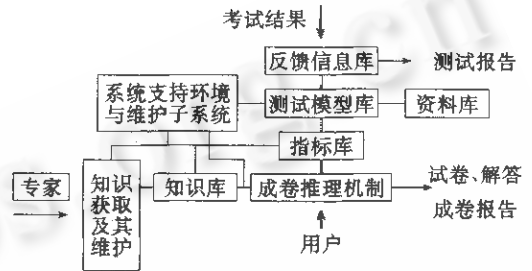


图1 系统结构

(b)为教务部门(使用成卷系统用户),提供安装、使用方法和专业术语的联机解释、求助功能.

2 成卷知识获取和表示

2.1 以“指标体系”为引导(indexes-directed)的知识获取与维护

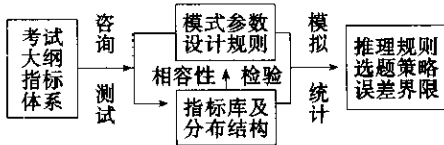


图2 知识确定次序图

在基于知识的计算机辅助命题及其分析外壳系统环境下,对任一课程,建立成卷系统的第一步是知识获取.知识包括:指标体系、试题指标值基本模式及经验规则,其中的指标体系是关键,它决定了指标库结构,基本模式及需要哪些命题经验规则;从成卷角度看,指标体系决定了成卷要求信息格式及成卷质量评判标准.因此,知识获取必须是以指标体系为引导的.成卷中的知识间的确定次序如图2所示.

以指标体系为引导的.成卷中的知识间的确定次序如图2所示.

2.2 成卷知识的层次结构和表示

为实现系统结构化,便于知识管理和维护,我们采用层次结构表示成卷知识,如图3.

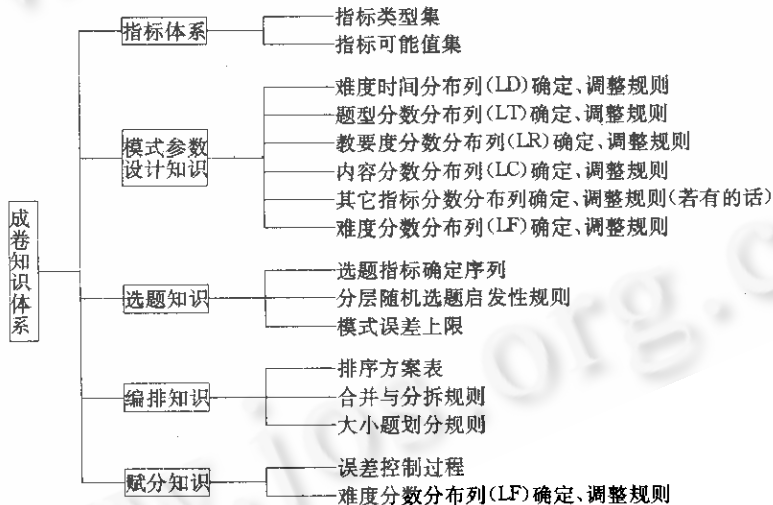


图3 成卷知识的层次结构

(1) 指标体系的建立

指标体系由指标集和各指标的可能值组成,以文本(.TXT)格式存放.据分析,试题指标可分为“共性”指标和特殊指标.“共性”指标是根据一般考试学原理,为达到考试科学性而设计的试题指标,它们是:难度、内容、教学要求度、时间值、题型、区分度;它们是建立题库的必要的的基本指标,统一的格式,但其可能值由课程定义,“特殊”指标是除共性指标之外的试题指标,含义和个数均由该课程专家而定.

(2) 指标库的建立

指标体系确定了试题卡的格式.对每道试题,在卡上由专家群或经过测试统计而得的值再加上其它试题信息,输入后,便构成了指标库,以关系数据库(.DBF)格式存放.因为设计

试卷参数设计是根据试题指标而进行的,成卷质量将直接依赖于试题指标的准确性和指标库的规模及分布结构.

(3) 试卷模式确定规则库的建立

试卷模式确定规则库存储该课程考试的构成基本模式的分布表组成(IF...THEN...格式)分布表的个数,将取决于指标体系.如果该课程的指标体系中还有其它指标用来刻划整卷性能,还可以建立该指标与分数的分配比例确定规则.

例如,对于试题难度指标,模式库中有一难度分数分布列库格式为:

$$\text{IF } (P, \sigma) \text{ THEN } \{(d_{i1}, d_{i2}, \dots, d_{iu}) \mid (i=1, 2, \dots, n)\}$$

其中 P, σ 分别为百分制下的平均分和根方差; d_{ij} 为 j 难度档的分数分布列 ($j=1, 2, \dots, u$), n 为可选的分布列的个数.

(4) 选题指标次序的确定

对给定的题库结构和基本试卷模式,对选题的所有可能的选题策略,用模拟方法,统计模式实现成功率,得到其最高策略.例如,某课程成卷系统的某一模式:

LD: (20.0, 70.0, 10.0),

LC: (20.0, 10.0, 15.0, 10.0, 0.0, 5.0, 10.0, 5.0, 5.0, 5.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0),

LT: (15.0, 75.0, 5.0, 5.0, 0.0),

LR: (0.0, 5.5, 80.0, 5.0, 0.0),

运行 11000 次,得最佳策略为: $D \rightarrow C \rightarrow T \rightarrow R$, 模式实现成功率为 92.5%.

(5) 成卷规则库的建立

成卷规则包括模式修正规则、模式检验规则、随机选题规则和试卷编排规则.

(a) 模式修正规则: 当前试卷模式必须与当前的用户要求结合. 运用模式修正规则, 得到今次考试的比例分布. 主要是比例分配的调整、转移、放缩规则.

(b) 模式检验规则: 当前试卷模式必须与题库结构相容(即能实现). 我们建立了相容性检验的数学模型^[4].

(c) 随机选题规则: 在确定了题库结构后, 选题时为了满足随机性要求的随机策略及提高搜索速度, 必须提供按指标确定次序进行的分层随机选题启发性知识.

例如: 假如任务日程表为: 难度 $D \rightarrow$ 内容 $C \rightarrow$ 题型 $T \rightarrow$ 教学要求度 R . 若当前选难度档为 D , 内容指标确定要考虑的因素有: ①哪些内容可以选(即: 有题, 用户有要求); ②人类专家更习惯(趋向)在哪些内容选. 此时, 应该用到规则:

$$\text{IF } D=k \text{ THEN } (p_{k1}, p_{k2}, \dots, p_{kv})$$

其中, p_{ki} 为选择第 i 内容的概率, $p_{k1} + p_{k2} + \dots + p_{kv} = 1$. 由此概率分布随机选得内容为 C_j , 为选题型, 运用下一层规则:

$$\text{IF } D=k, C=j \text{ THEN } (q_{kj1}, q_{kj2}, \dots, q_{kjw})$$

其中, q_{kji} 为选择第 i 题型的概率, $q_{kj1} + q_{kj2} + \dots + q_{kjw} = 1$.

(d) 试卷编排规则: 选到一组试题集后, 确定在试卷中的先后次序; 由于题库中的试题, 无论是难题还是易题, 均是以单题形式存放, 试卷编排时要将一些小题合并成大题, 更符合考试习惯或规律, 有利于考试信度和评卷等. 所以, 编排规则包括大小题划分规则、排序规则

和合并规则(为判定表形式). 例如,“概率论与数理统计成卷系统”中,有如下大小题划分规则:

R126:IF $D \geq 4$ AND $T = *$ AND $S > 0$ AND $F > 0$ THEN SINGLE_ITEM = 1

R127:IF $D = 3$ AND $T = 4$ AND $S \geq 10$ AND $F > 0$ THEN SINGLE_ITEM = 1

R126 表示了当难度档为“较难”,其它条件为任意时,则该题为大题单题处理;R127 表示了当难度档为“中等”的论证题,答题时间 10 分钟以上,则该题为太题单题处理.

3 GITB shell 的试卷设计的推理机制

1. GITB shell 的试卷设计的推理机制用 Borland C++ 3.1 for Windows 编程,具有如下特点:

(1)能适应和感知用户的各种变化的要求,支持用户的合理的人工干预.

(2)多目标性. 成卷是在有限资源(即试题库)下的多目标设计,由于不同的设计目标相互竞争,需要考虑大量的折衷和目标的冲突处理.

(3)解空间大. 即对于相同的命题要求,一般可设计出几个不同形式等效的试卷模式.

(4)随机性. 这是因为:①体现机器成卷命题的特色,避免教师命题的主观性;②以保证题库中符合条件的题目都有机会被选中,充分利用题库的资源. 我们建立多层次随机选题数学模型,保证系统在产生试卷模式及选题过程具有尽可能大的随机性^[4].

(5)近似性. 试题的各指标值无论是通过测试,还是专家综合评定,都带有一定的近似性和对测试样本的相对性;另外,在选题组卷时,由于受题库中题的分布状况的影响,有时不能完全按模式实现,虽然建立了误差控制策略数学模型,仍会产生一定的误差.

2. 成卷过程是将一组命题要求转化为寻找满足该组要求的题库中试题组成的试题子集的过程.

命题要求可分为 4 类:(1)今次考试的宏观要求,如课程名、考试目的、范围、考生对象(层次),能力侧重点,答卷时间,平均分,及格率,要否副卷等;(2)第一类命题要求中的基本信息及部分有关该课程考试的试题指标比例要求,如:在考试范围内,某些内容的分数比例;(3)有关该课程今次考试模式信息;(4)今次考试的各道试题的详细指标(包括分数)要求值. 系统能对其中的矛盾要求给予警示.

对于第四类命题要求,成卷系统的任务是指标值的匹配问题,只要试题结构分布合理,有良好的用户人工干预界面和随机性,就可以完成. 而对于前三类要求,成卷系统首先要解决的问题是如何模拟人类专家的命题思维过程.

教学专家在长期的教学实践中,积累了丰富的命题经验,形成了对某种考试的模板(即原型),对于常规性考试,在命题工作的初始阶段,首先使用以往的跟当前相关的经验模板. 当没有相关的模板与之匹配时,才从一般类似的情况推理得出新的设计方案.

根据基于“原型”的知识表示方法,以及成卷任务的特点,由一组比例关系来描述的试卷整体性能要求,我们称之为试卷模式(exam-paper pattern). 有了整卷的基本模式,通过修正、优化,得到当前成卷模式. 下面的任务是如何运用一定的搜索规则,从试题库中挑选试题,满足多种设计约束要求.

我们建立面向对象的程序设计(OOP)模块结构^[5],如图4.将试卷模式作为主类,当前试卷模式作为主对象,将成卷任务划分成几个相对独立的子任务:

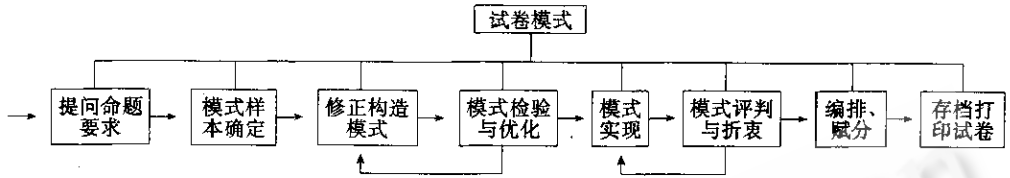


图4 成卷子系统的工作流程

①模式选择(pattern_selecting()):从命题要求由分析器进行类比分析、匹配,选择到一个参考试卷模式,即由考生层次、考试目的、能力侧重点、考试范围等,去匹配出一条条模式分布列(为确定LD,需先确定P、 σ),构成当前试卷模式;例如:“金属工艺学成卷系统”的一组命题要求:

考生层次:01(机类学生) 考试目的:02(总结性)
能力侧重点:1(基本技巧) 考试范围:01(全部)

匹配得一个模式:

LD:(20.0,70.0,0.0)

LC:(8.0,4.0,4.0,4.0,10.0,20.0,4.0,4.0,4.0,10.0,4.0,4.0,5.0,5.0,5.0,5.0)

LT:(15.0,75.0,5.0,5.0)

LR:(0.0,0.0,80.0,5.0,0.0)

②模式修正(pattern_adjusting()):从参考模式出发,结合当前具体要求(如命题时间TOTAL)、特殊要求(如第二、三类用户中可指定LC中的一些分量),运用调整模式规则,得到一个当前试卷模式.上例中,若命题时间TOTAL=120(分钟),LC中c1=15.0(分钟),c2=5.0(分钟),得

LC:(15.0,5.0,4.5,4.5,11.4,22.8,4.5,4.5,4.5,11.4,4.5,4.5,5.7,5.7,5.7,5.7,5.8),其余分布列略.

③模式检验与优化(pattern_test_and_optimizing()):从一组命题要求中,由①②可得若干模式,对每一模式,计算其与题库结构的相容度,选其通过检验且相容度最大者,作为当前模式,转④.若所有被选模式均通不过检验,转②;

④模式实现(pattern_realizing()):根据当前试卷模式,从试题指标库中,运用选题规则和模型选出一组试题,逐步逼近模式,最后得到一个由这些试题指标描述的实际试卷模式.每选一道题,需按“选题指标确定序列”确定其指标要求:

step1:对第一个指标,按其“当前时间分布列与题库结构的相容度”的“小优先”原则确定;

step2:其它指标,运用分层随机启发式规则及当前分数分布列确定;

step3:按难度时间分布列及其误差界限,确定时间指标;

step4:找题;

step5:找题失败时,运用下列冲突求解策略解决:

(1)约束折衷,按折衷规则,再找题.例如,在“高等数学成卷系统”的指标确定步骤中,出

现了从能力设计考虑,希望选“证明题”,从结构效度考虑,希望选“不定积分”,而“不定积分”部分没有合适的证明题,这时可折衷选“不定积分概念题”;

(2)逐步放宽约束,再找题;

(3)改变选题指标确定次序,再找题;

(4)用户干预.

step6:找题成功时,改变当前试卷模式;

⑤模式评判(pattern_judging()):模式评判是检验当前试卷模式分量卷模是否达到误差界限,若是,转⑥,否则转④.

⑥赋分(score_assigning()):运用赋分误差控制方法和难度分数分布列(LF)确定、调整.整卷得分率误差 ΔE 控制在:

$$|\Delta E| \leq \max\{|(a_{i+1} - a_i)\Delta F_i|\} \quad 0 < i < u$$

其中 a_i 为难度档为 i 的得分率, ΔF_i 为难度档 i 的题的赋分误差.

⑦编排(merge_and_sort())分为三步骤:

step1:大小题划分

step2:小题合并、分拆及排序,产生新大题

step3:大题排序

另外,以上各步均有手工干预窗口;为控制选题重复率、测试和系统维护的目的,成卷信息存档是必不可少的.

4 结束语

本文阐述通用智能成卷系统外壳 GITB shell 的总体设计方案及技术.我们吸取了多年来这方面的各学科成卷系统研究经验,如高等数学、大学物理、大学化学和理论力学等试题库系统,充分考虑到各课程的特点,采用基于原型的构模方法、面向对象程序设计的开发方法和最合适的计算机软件开发环境.该系统已近完成,以其为工具制作的“全国工科金属工艺学成卷系统”和“全国工科概率论与数理统计成卷系统”已基本制成.实践表明,本系统是有效的、实用的,达到了预期目标.至于其中的数学模型、算法设计等问题,将在另文叙述.由于通用性和智能性的要求,该课题的难度和复杂性大大加强,许多问题正在进一步完善,如考试测试与分析模型、知识获取自动化等.

致谢 感谢 GITB shell 项目组的其他所有成员的大力支持.

参考文献

- 1 国家教委考试中心.题库建设理论与实践.光明日报出版社.
- 2 蒋康明.物理题库智能组卷系统 PBICS 的研究与实现[硕士论文].华南理工大学,1990.
- 3 陈永乐.一个成卷专家系统的设计与实现[硕士论文].华南理工大学,1989.
- 4 林复华.智能成卷系统中的数学模型.算法及其应用[硕士论文].华南理工大学,1992.
- 5 Leung K S, Wong M H. An expert-system shell using structured knowledge — an object-oriented approach. IEEE Computer, 1990, 23(3): 38-47.

DESIGN SCHEME OF KNOWLEDGE — BASED COMPUTER — AIDED EXAM — PAPER FORMATION SYSTEM

Lin Fuhua Chen Guangzhong Hong Chaoxing Wang Guoqiang

(*Department of Applied Mathematics, South China University of Technology, Guangzhou 510641*)

Abstract In this paper, the authors first discuss the structure and functions of the knowledge — based computer — aided exam — paper formation and analysis system shell, and then present the tactics of “indexes — directed” knowledge acquisition and the structure and representation methods of the knowledge base, and then introduce the inference mechanism of exam — paper design as well as the “exam — paper pattern — oriented” solving model, the layered random selecting and conflict — solving rules.

Key words Computer — aided exam — paper formation, knowledge acquisition, expert system, system shell.