

题库系统智能成卷理论和组卷方法研究

王友仁, 张 晔, 施玉霞, 姚 睿

(南京航空航天大学自动化学院 南京 210016)

【摘要】分析了组卷目标要求,建立了智能组卷系统问题求解的理论模型。结合组卷问题求解领域知识与专家选题思想,提出了一种基于自适应混合进化机制的智能组卷方法,采用了基于题型分段编码、段内进化算子、误差加权适应度评估方法。通过实例分析,实验结果表明所提出的新组卷方法具有组卷成功率高和组卷速度快的优点。

关键词 人工智能专家系统; 试卷自动生成; 理论模型; 进化算法
中图分类号 TP18 文献标识码 A

Theoretical Model and Method of Intelligent Test Paper Generation System from Item Pool

WANG You-ren, ZHANG Zhai, SHI Yu-xia, YAO Rui

(College of Automation and Engineering, Nanjing University of Aeronautics and Astronautics Nanjing 210016)

Abstract Based on the analysis of the characteristics of the test paper, a theoretical model of intelligent test paper generation system is established. Then the method of test paper auto-generation based on Adaptive Hybrid Evolutionary Mechanism (AHM) is proposed. The experimental results show that new method can solve the problem of test paper auto-generation effectively and efficiently.

Key words artificial intelligent system; automatic generation of test paper; theoretical model; evolutionary algorithm

计算机辅助教学(Computer Assistant Instruction, CAI)已经成为当前高等教育的重要组成部分,电子题库技术与自动组卷系统是CAI的重要研究方向。自动组卷是实现无纸化考试、考试标准化与个性化、网络化学习与远程教育在线测试等的基础。自动组卷是把人工智能技术与人类教育专家的知识和经验相结合,由计算机来自动设计试卷内容,且使计算机生成的试卷能达到专家级水平。自动组卷系统研究的主要难题是如何保证自动生成的试卷既能满足用户的需要,又能保证其科学性、随机性、客观性,以及远程网络交互式环境下的实时性。因此,高效、简单通用、鲁棒性好的智能组卷算法是实现自动组卷的关键技术和研究热点之一。

目前智能组卷算法主要有随机抽取法、回溯试探法、遗传算法^[1-3],这些算法常因约束条件的局部满足而导致组卷失败,或选取试题缺乏随机性,组卷过程所需时间也较长,无法满足实时应用。基于此,本文提出了一种基于自适应混合进化机制(Adaptive Hybrid Evolutionary Mechanism, AHM)的智能组卷方法。

1 题库系统智能成卷的问题描述

计算机根据用户的要求,搜索电子题库中属性参数相匹配的试题,自动生成让用户满意的试卷。智能组卷系统基础理论主要涉及试卷质量评价与组卷目标要求、问题求解的理论模型和组卷方法。

1.1 试卷指标与组卷目标

要使计算机能自动生成试卷,首先涉及试卷质量评价与组卷目标问题。根据教育测量与统计学理论,传统试卷质量评价指标主要以试卷的难度、区分度、信度和效度等作为衡量的标准^[1-3]。试卷的难度是考试试卷难易程度的评价指标;试卷的区分度是试卷能够区分鉴别考生不同知识能力水平的衡量指标;试卷的

收稿日期:2004-05-17

基金项目:江苏省教育厅基金资助项目(P0262-JW)

作者简介:王友仁(1963-),男,博士,教授,主要从事计算机CAI系统、电子设备测试诊断与自修复、仿生硬件等方面的研究。

信度是指考试分数能真实反映学生水平的可靠性指标；试卷的效度是指考试能有效地对教学效果进行准确测量评价的性能指标。试卷的信度和效度一般只能定性评估。试卷难度的一般计算公式为：

$$D = \frac{1}{P} \sum_{i=1}^M d_i p_i = 1 - \frac{\bar{Y}}{P} \quad (1)$$

式中 P 为试卷的满分值； d_i 为第 i 道试题难度值； p_i 为第 i 道试题分值； M 为试卷的题目总数； \bar{Y} 为该份试卷全部考生的平均得分值。各道试题难度值 d_i 需要大量数据来统计分析，无法精确测量。

试卷区分度

$$F = \frac{1}{P} \sum_{i=1}^M F_i p_i = a\delta + b \quad (2)$$

式中 F_i 为第 i 道试题的区分度； a 、 b 为常数(由经验确定)； δ 为试卷考试成绩的标准差。试卷区分度与各试题的区分度 F_i 有关，而 F_i 难以精确计算，其初值常取经验值，再根据考试情况进行调整。

传统的试卷质量评价指标是基于试题属性值，而统计测量与专家经验都难以获得精确属性值。因此，必须结合定量与定性方法，综合考虑到一般试卷质量指标、教学要求、特定用户需要，来确定合适的试题属性值，建立科学的试卷评价准则，形成智能组卷目标要求。在自动组卷系统中，试题分数、估计用时为定量表示，试题所属知识点、试题类型、知识掌握要求用符号表示，试题难度、区分度采用模糊分级表示。组卷目标应体现试卷质量评价指标与教学要求，所设计的计算机组卷目标为：试卷总分、章节分值、题型分值、试卷难度、考试时间、知识点满足、能力层次满足、试卷区分度、试卷形式。

1.2 智能组卷的理论模型

计算机自动组卷过程为在一定题量的试题库中搜索满足组卷目标要求的一组属性试题组合。根据组卷目标分析来确定每道试题的主要属性代码，对于试题库中全部试题可以预先建立目标状态矩阵 A ：

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,C} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,C} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,C} \end{bmatrix} \quad (3)$$

式中 N 为试题库中全部试题总数； C 为每道试题的属性代码项数(即对应生成试卷时最多的组卷目标要求，这里 $C=9$)。矩阵 A 中每行代表一道试题及属性，每列代表试题库中全部试题的某个属性。

智能成卷问题求解过程为：根据用户的组卷要求，在矩阵 A 中寻找满足组卷目标要求的行组合，即搜索确定矩阵 A 中全部行状态的多变量优化组合 (K_1, K_2, \dots, K_N) 。其中 K_i 为第 i 行的状态变量(即第 i 道题)，若 $K_i=1$ 表示第 i 行被选中，而 $K_i=0$ 表示第 i 行没有被选中。由此，可构造如下数学模型：

$$(1) \text{ 试卷总分: } P \text{ 为试卷总分(默认值100分), } a_{i,1} \text{ 对应第 } i \text{ 道题的参考分值。则有: } \sum_{i=1}^N K_i a_{i,1} = P。$$

(2) 各章节分值：用 s 表示20个章节的序号， Y_{Q_s} 为用户要求的第 s 章节的分值。 $a_{i,2}$ 为第 i 道题的章节序号值，当 $s = a_{i,2}$ 时，有 $h_{u_{i,s}} = 1$ ，否则 $h_{u_{i,s}} = 0$ 。各章节试题分值和等于试卷总分。则有：

$$\sum_{i=1}^N K_i h_{u_{i,s}} a_{i,1} = Y_{Q_s} \quad s = 1, 2, \dots, 20 \quad (4)$$

(3) 各题型分值：试题共分成7种类型，分别为：A选择填空、B问答题、C改错题、D证明题、E分析判断题、F计算题、G综合题。各类型题分数之和等于试卷总分。则有：

$$\sum_{i=1}^N K_i h_{h_{i,m}} a_{i,1} = Y_{W_m} \quad m = 1, 2, \dots, 7 \quad (5)$$

式中 m 为试题库中所有题型的代码序号； Y_{W_m} 为用户要求的第 m 种类型试题的分值。 $a_{i,3}$ 为第 i 道题目类型属性代码序号，当 $m = a_{i,3}$ 时，有 $h_{h_{i,m}} = 1$ ，否则 $h_{h_{i,m}} = 0$ 。

$$(4) \text{ 试卷难度: 设 } D \text{ 为用户指定的试卷难度值, } a_{i,4} \text{ 为第 } i \text{ 道题的难度值。则有: } \sum_{i=1}^N K_i a_{i,1} a_{i,4} = PD。$$

$$(5) \text{ 考试时间: 设 } a_{i,5} \text{ 为第 } i \text{ 道题的预计答题时间, } T \text{ 为用户要求的考试时间(默认值为120 min)。则:}$$

$$\sum_{i=1}^N K_i a_{i,5} = T \quad (6)$$

(6) 知识点满足: 用户可指定试卷包含某知识点属性的试题与分值大小。则有:

$$\sum_{i=1}^N K_i h_{g_{i,n}} = Y_{U_n} \quad n=1,2,\dots,30 \quad (7)$$

式中 n 为试题库中所有30个知识点的代码序号, Y_{U_n} 为用户要求的第 n 知识点的分值。 $a_{i,6}$ 为第 i 道题的知识点属性代码序号值, 当 $n=a_{i,6}$ 时, 有 $h_{g_{i,n}}=1$, 否则 $h_{g_{i,n}}=0$ 。

(7) 能力层次满足: 学生掌握教学内容可分为了解、理解、熟练掌握和灵活运用4个层次。设 t 表示能力层次属性代码序号, Y_{M_t} 为要求的能力层次试题分值。 $a_{i,7}$ 为第 i 道题的能力层次属性代码序号值, 当 $t=a_{i,7}$

时, 有 $h_{t,i}=1$, 否则 $h_{t,i}=0$ 。则有: $\sum_{i=1}^N K_i h_{t,i} a_{i,1} = Y_{M_t}$, $t=1,2,3,4$ 。

(8) 试卷区分度: 设 F 为用户要求的试卷区分度, $a_{i,8}$ 为第 i 道题的区分度。则有: $\sum_{i=1}^N K_i a_{i,8} = PF$ 。

(9) 试卷形式: 分为客观形式和主观形式两种。前者支持计算题的标准化解答, 后者需要人工解答。设 b 表示试题试卷形式序号, R_b 为用户指定试卷形式。 $a_{i,9}$ 为第 i 道题的试卷形式属性代码。则有:

$$\prod_{i=1}^N ((a_{i,9}|_{K_i=1}) \subset R_b) = 1 \quad b=0,1 \quad (8)$$

由此可见, 智能组卷属于多目标组合优化问题, 问题求解精度(即组卷目标的满足程度)越高表明试卷质量越好。这里, 用户也可根据实际需要来指定组卷目标的数量。

2 组卷方法

遗传算法可以用于求解智能组卷问题^[1-3], 但基本遗传算法存在容易早收敛^[4-5]、进化后期搜索效率较低的缺点, 故使用效果并不令人满意。为此, 根据组卷问题中试卷构成知识、专家设计试卷时选题思路, 提出一种基于自适应多点变异混合算法的智能组卷方法。新算法的构造要素设计如下:

1) 试题编码: 设题库中共有 N 道题、7种题型(各题型分别有 $N_1 \sim N_7$ 道题)。根据题库中同类型试题分值多数相同的特点, 采用按试题类型进行分段编码方案。首先对各类型试题进行独立编码, 再合并成 N 位二进制代码, 即染色体代码为 $(J_{1,1}, J_{2,1}, \dots, J_{N_1,1}), \dots, (J_{1,7}, J_{2,7}, \dots, J_{N_7,7})$, 若 $J_{i,j}=1$ 表示第 i 类题型中第 j 道题被选中, $J_{i,j}=0$ 表示没被选中。设组卷要求的各题型试题数量分别为 M_1, M_2, \dots, M_7 , 则应该满足:

$\sum_{i=1}^{N_k} J_{i,k} = M_k$, $k=1,2,\dots,7$ 。这种分段编码策略可有效缩小解的搜索空间, 简化了问题的求解。

2) 适应度函数: 构造染色体的误差函数为: $E = \sum_{k=1}^9 \alpha_k e_k$ 。

式中, $e_k=0 \sim 1$ 为当前代染色体满足第 k 项组卷目标要求的归一化相对误差模, $0 < \alpha_k < 1$ 为权重系数。对于试卷总分值、各题型分值、时间、难度等组卷目标取较大的 α_k , 章节内容、教学内容掌握能力、知识点等目标项所对应 α_k 值为中等, 而试卷形式、区分度等目标项对应较小的 α_k 值。染色体的适应度函数定义为: $F_{\text{fit}} = (E + 0.01)^{-1}$ 。

3) 遗传算子设计: (1) 分段多点变异操作: 根据专家设计试卷的选题思想, 对染色体中各个独立编码段依次进行段内自适应多点基本位变异操作。基本位变异操作是对段内编码串以概率1随机指定的某一位基因座上的基因进行变异, 在进行段内某基因位变异后, 再在该变异位所在段内, 随机向后或向前搜索寻找与该变异位基因值相反的基因座(将基因位值取反)。染色体中各段内多点基本位变异数为:

$$M_{\text{BB}_k} = \omega_k (F_{\text{fit}_{\text{max}}} - F_{\text{fit}}) / F_{\text{fit}_{\text{max}}} \quad k=1,2,\dots,7 \quad (9)$$

式中 ω_k 为第 k 段代码的最大变异位数, 可取 $\omega_k = (0.001 \sim 0.01)N_k$; $F_{\text{fit}_{\text{max}}}$ 为最大适应度; F_{fit} 为当前代染色体适应度。若变异位数小于1, 则以概率0.5在段内随机确定一个变异点。选择更新策略: 根据子代个体的适应度来确定分段变异后的子代个体能否代替父代个体。若子代个体的适应度比父代个体的大, 则子代替换父代。否则, 按照一定选择概率来确定替换。为了使搜索过程避免陷入局部最小、提高全局搜索能力, 结合模拟退火算法的思想, 选择概率为:

$$p_{ss} = \rho / \{1 + \exp[(F_{\text{fit}} - F'_{\text{fit}}) / (\lambda^j F_{\text{fit,max}})]\} \quad (10)$$

式中 $\rho = 0.5 \sim 1.0$ 是常数; $\lambda = 0.65 \sim 0.95$ 为退温速率; j 是进化代数; F'_{fit} 为子代个体适应度。

进化过程结束条件:(1) $F_{\text{fit}} \geq 98$; (2) $F_{\text{fit}} \geq 95$, 且在连续6代内适应度改善小于0.2; (3) 进化到最大代数。若 $F_{\text{fit}} < 95$, 则认为搜索失败。

开始时随机生成各个独立代码段代码, 且保证初始染色体已满足各题型分值与总分值要求, 这样可减少组卷问题求解的总约束项。新算法实现步骤为:(1) 初始染色体生成;(2) 分段多点变异操作;(3) 选择更新操作;(4) 计算适应度;(5) 判断满足进化结束? 若不满足, 则返回(2)。否则结束进化过程。

3 应用实例

本文以模拟电子技术题库系统为例, 来分析智能成卷系统的实现与实验结果。电子题库中每道试题的存放信息包括属性代码及题序、试题内容、参考答案、考试反馈信息(答题情况与选题次数)。试题属性代码由16位组成, 例如某属性代码为0402110008423006, 其含义为: 第4章中第2个知识点、选择题(1)、分值10、客观题(0)、8 min、区分度为4、难度中等(2)、题序是3 006。

表1 典型试卷的试题难度分布

试题难度等级	E	D	C	B	A
难度等级分值	15	28	42	10	5

表2 典型试卷的各章分值分布

章号	1	2	3	4	5	6	7	8	9	10
分值	5	20	12	20	16	10	11	0	6	0

需要确定的组卷要求有: 试卷总分、各题型分值、各章节分值、试卷难度、考试时间、高级设置项(各章节知识点、全卷区分度、知识能力掌握、试卷形式), 用户可逐项设定要求或使用默认值。试卷难度分成A、B、C、D、E5个等级, 分别为难A(难度系数 n_d 为0.6~1.0)、比较难B(n_d 为0.35~0.6)、中等C(n_d 为0.2~0.35)、比较容易D(n_d 为0.1~0.2)、容易E($n_d < 0.1$)。

新算法AHM的参数为: $\omega_1 = 5$, $\omega_k = 2$, $\rho = 0.6$, $\lambda = 0.9$ 。基本遗传算法(Simplex Genetic Algorithm, SGA)的参数为: 种群规模60, 分段单点交叉概率0.5,

表3 不同组卷算法的组卷结果对比

算 法	平均收敛代数	平均收敛时间/s	组卷成功率/(%)
AHEM	1 150	17.6	96.5
SGA	485	151.8	85.5

分段均匀变异概率0.001, 采用比例选择策略。试题数量为1 500, 计算机为P4-1.7G(内存512 M), 程序采用VB6.0编程。表1、表2为自动组卷典型试卷指标, 表3为200次组卷结果统计。由实验结果可知, 智能

组卷算法能获得满足组卷要求的试卷, 新算法在组卷成功率与组卷时间上要明显优于SGA。

4 结 论

智能组卷系统具有成卷快、客观性好、难易可度量、能同时生成多份试卷的特点, 本文根据组卷问题特点与专家思想, 提出的基于自适应进化机制的智能组卷算法提高了组卷成功率和成卷速度。

参 考 文 献

- [1] 王宇颖, 侯 爽, 郭茂祖. 题库系统试卷自动生成算法研究[J]. 哈尔滨工业大学学报, 2003, 35(3): 342-346.
- [2] 魏 平, 张 元. 一种求解组卷问题的遗传算法[J]. 宁波大学学报(理工版), 2002, 15(2): 47-50.
- [3] 全惠云. 基于遗传算法的题库智能组卷系统研究[J]. 武汉大学学报(自然科学版), 1999, 45(5): 758-760.
- [4] Galante M. Genetic algorithms as an approach to optimize real-world trusses[J]. International Journal for Numerical Methods in Engineering, 1996, 39(1): 361-382.
- [5] Leung Y W, Wang Y P. Multi-objective programming using uniform design and genetic algorithm[J]. IEEE Trans Systems, Man and Cybernetics, Part C: Applications and Review, 2000, 30(3): 293-304.

编 辑 漆 蓉