

基于自适应遗传算法的智能组卷研究

董敏, 霍剑青, 王晓蒲

(中国科学技术大学 人工智能与计算机应用实验室, 安徽 合肥 230026)

摘要: 计算机辅助教学 CAI (Computer Assisted Instruction) 的一个重要应用是计算机辅助测验 CBT (Computer-Based Testing). 智能组卷是 CBT 的基础, 组卷中关键是解决约束优化问题. 在研究现代教育测试理论与计算机辅助测验 CBT 的基础上, 提出一种解决计算机组卷中约束优化问题的方法——自适应遗传算法. 该方法有效地解决了基于 IRT 的智能组卷问题, 为解决约束优化问题提供一种新的有效途径, 具有较好的性能和实用性.

关键词: 项目反应理论 IRT; 自适应遗传算法 AGA; 约束优化问题

中图分类号: TP18

文献标识码: A

文章编号: 1000-1220(2004) 01-0082-04

Research on Computerized Test Construction Based on Adaptive Genetic Algorithms

DONG Min, HUO Jian-qing, WANG Xiao-pu

¹(Artificial Intelligence & Computer Application Laboratory, University of Science & Technology of China, Hefei 230026, China)

Abstract: One of the important applications on Computer Assisted Instruction (CAI) is Computer-based Testing. Computer Test Construction is the basis of CBT, which key problem is to solve constraint optimization problems. On the research of modern test theory and CBT, an adaptive genetic algorithm is given here to solve the problems of computer test construction based on IRT. The results indicated that AGA could solve the problems effectively, which gives us a new way to solve constraint optimization problems with good performance and practicability.

Key words: item response theory (IRT); adaptive genetic algorithms (AGA); constraint optimization problems

1 引言

测验是评价教学的一个重要环节, 随着计算机的广泛应用, 计算机自动组卷代替了手工组卷, 节省了人力, 提高了组卷的有效性. 目前国内的组卷系统所采用的方法都是随机抽取试题组卷或回溯法生成试卷. 在确定试卷目标要求时, 完全以经典测量理论 CTT (Classical Test Theory) 为指导^[1], 这样, 所采用的计量学指标值不仅依赖于被测样本, 而且无法确定单题性能与整卷功能的关系. 这种先天缺陷就使选题工作具有很大的盲目性, 并且对组卷要求比较严格、组卷指标比较复杂的情况不能很好地满足要求.

本文是在研究现代教育测试理论模型与计算机辅助测验 CBT 的基础上, 提出一种解决计算机组卷中约束优化问题的方法——自适应遗传算法. 通过模拟实验, 结果表明该方法有效解决了智能组卷中的约束优化问题, 具有很好的性能和实用性.

2 现代测量理论—IRT

项目反应理论 (Item Response Theory) 简称 IRT, 是在克服了经典测量理论 CTT (Classical Test Theory) 的各种局限性而发展起来的一种全新的测量学理论^[2,3]. 目前应用较广的 IRT 模型是单维逻辑斯蒂模型, 单维三参数逻辑斯蒂模型的函数表达式为:

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-D \cdot a_i(\theta - b_i)}} \quad (1)$$

其中, $P_i(\theta)$ 是项目特征函数 ICF (Item Characteristic Function), 表示能力参数为 θ 的考生, 对测验项目 (题目) $i = 1 \dots I$ 能做出正确的反应的概率. a_i 是项目 i 的区分度参数, b_i 是项目 i 难度参数, c_i 是项目 i 的机遇参数 (猜测参数). D 为常数, 通常取 $D = 1.704$. 三参数逻辑斯蒂模型的项目特性曲线如图 1 所示.

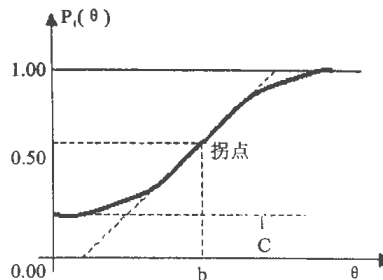


图 1 逻辑斯蒂模型项目反应特征曲线 ICC

Fig. 1 Item characteristic curve of logistic model

根据数理统计中的 Fisher 信息函数定义, 在项目反应模型下, 能力参数未定的被试在 n 个测验项目上的信息函数为^[2]

$$I(\theta) = Var(\hat{\theta}|\theta^{-1}) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{[P_i(\theta)]^2}{P_i(\theta)[1 - P_i(\theta)]} \quad (2)$$

$I(\theta)$ 与 $I_i(\theta)$ 分别表示测验信息函数(TIF)和项目信息函数(IIF)。 $I(\theta)$ 可用来衡量测验测量考生能力的准确性。因此,组卷可通过测验信息函数来实现,即通过在某些特定的能力点 θ 与指定的测验信息函数 TIF 相匹配而进行基于 IRT 的组卷,这样能主动按照反映测验目的性质要求的测验目标信息函数 TIF 曲线来挑选题目,使测验编制过程显出清晰的科学逻辑。

3 智能组卷问题研究

3.1 组卷中的约束优化问题

组卷问题可被描述为一个约束的组合优化问题,该问题被定义为一个目标函数与一组约束条件的组合^[4]:

3.1.1 目标函数:基于 IRT 组卷的目标函数有多种形式,如:最大化 TIF 的相对形态,最小化测验长度,最大化在指定能力处的 TIF,最小化相对目标测验的绝对偏差和,等等。有时,目标函数是与一个目标测验相匹配,或者尽量接近某些指定能力处的 TIF。

3.1.2 测验要求:是测验选题的规则,包括在心理测量、测验内容、题目格式等属性上的约束限制,即在某能力点满足指定 TIF,设置测验中包含某类内容的题目数,以及限制测验中具有某种题目格式或题目类型的题目数等。同时包括项目间依赖性的约束限制,表现为不能同时出现的题目之间的互斥关系或必须同时出现的题目之间的包含关系。

本文将基于项目反应理论的组卷问题描述为一个最大化模型 I-MAX^[4],满足测验条件⁽⁵⁾ ⁽⁶⁾ ⁽⁷⁾ 的同时,匹配⁽⁴⁾ 中对目标测验信息函数指定的形状,最大化测验信息函数:

$$\text{maximize } y \quad (3)$$

$$s.t.: \\ y \leq \frac{\sum I_i(\theta) x_i}{r_k}, \forall k \quad (4)$$

$$\sum_i q_{in} x_i \leq Q_n, \forall n \quad (5)$$

$$C_m^l \leq \sum_i c_{im} x_i \leq C_m^u, \forall m \quad (6)$$

$$\sum_i l_{ip} x_i \leq L_p, \forall p \quad (7)$$

$$x_i = \begin{cases} 1, & \text{题目 } i \text{ 在测验中,} \\ 0, & \text{否则} \end{cases} \quad \forall i$$

其中变量 x_i 是决策变量,标识一个题目是否被选到测验中,若被选中其值为 1,否则为 0。 I_k 是能力 θ 点的项目信息函数值,而 r_k 是测验目标信息量。系数 q_{in} 是题目 i 具有属性 n 的参数值, Q_n 代表测验中的包含属性 n 的题目数; c_{im} 是题目 i 的类别决策值,即如果题目 i 属于类别 m 则其值为 1,否则为 0; C_m^l 与 C_m^u 代表测验中允许包含的某一类别 m 的题目数范围;而 l_{ip} 是题目间约束关系的系数, L_p 是约束边界值,⁽⁷⁾ 用于表示题目之间的互斥或包含关系。

3.2 自适应遗传算法描述

遗传算法 GA (Genetic Algorithm) 是模拟生物界自然选择和自然遗传机制的随机化搜索算法,由 John Holland 提出,利用简单的编码技术和“繁殖机制”来表现复杂的现象,适用于处理传统搜索方法难于解决的复杂和组合优化问题^[6]。

本文提出了基于 IRT 组卷的自适应遗传算法,利用遗传算法的基本思想,引用惩罚函数^[6]的方法自适应地调整适应度函数,不仅避免了遗传算法中经常出现的“早熟现象”,而且有效地解决了智能组卷中的约束优化问题,具有很好的性能和实用性。

自适应遗传算法流程图如图 2 所示,

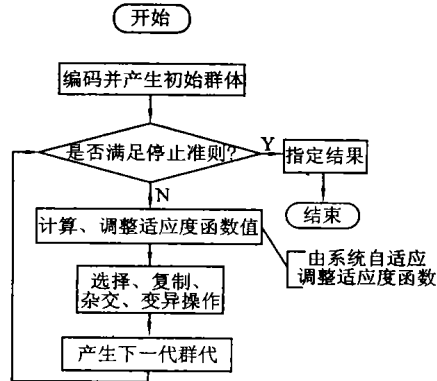


图 2 自适应遗传算法流程图
Fig.2 Flow chart of adaptive GA

在求解时,关键在于完成以下几个方面:

首先对优化问题的解进行编码,此处,我们称一个解的编码为一个染色体,组成编码的元素称为基因,编码的目的主要是用于优化问题解的表现形式和利于之后遗传算法的计算。

第二是适应函数的构造和应用。适应函数基本上依据优化问题的目标函数而定。当适应函数确定以后,自然选择规律是以适应函数值的大小决定的概率分布来确定哪些染色体适应生存,哪些被淘汰,生存下来的染色体组成种群,形成一个可以繁衍下一代的群体。

第三,染色体的结合。双亲的遗传基因结合是通过杂交 (crossover) 达到下一代的产生,新一代的产生是一个生殖过程,它产生了一个新解。最后是变异,新解产生过程中可能发生基因变异,变异使某些解的编码发生变化,使解有更大的遍历性。

最后是确定终止规则。一个最为简单的停止规则是给定一个最大的遗传代数,算法迭代代数达到这个值时停止;另一类终止规则是给定问题的一个下界的计算方法,当进化中达到要求的偏差度时,算法停止;还可以检测算法再进化是否还会改变解的性能,如果不能改变,则算法停止。

3.2.1 编码方式 假设题库中总共有 I 道题,用一个 I 位的二进制串来表示,其中 $x_i = \begin{cases} 1 & \text{当第 } i \text{ 题被选中} \\ 0 & \text{当第 } i \text{ 题未被选中} \end{cases}$

初始化的串群体,即随机生成 N 个串的群体;在串群体中,串长度都是相同的,群体大小根据需要给出,一般取个体编码长度数的一个线性倍数,如 I 和 $2I$ 之间的一个确定数,选择分布均匀的 0-1 二进制串能使算法更加有效。

3.2.2 自适应的适应函数(Fitness) 在遗传算法中,类似生物界自然选择过程和遗传、变异过程,出现在群体中个体的适应值引导算法把一个群体变换到一个新的群体。

适应函数和目标函数是紧密相关的;在遗传算法中,最大化适应函数,它的定义必须满足下面两个条件:首先,适应函数是生成个体的函数;其次,当目标函数达到最优时,适应函数达到最大值。

通过分析组卷的约束优化问题(3)~(7),算法的适应度函数以下面的形式给出:

$$f(x) = y - g(x) \quad (8)$$

$$g(x) = \lambda \sum_n h(\sum_i q_{in} x_i - Q_n) + \mu \sum_m h(C_m^l - \sum_i c_{im} x_i) + \mu \sum_m h(\sum_i c_{im} x_i - C_m^u) \quad (9)$$

$$h(u) = \begin{cases} u, & u > 0, \\ 0, & u \leq 0 \end{cases} \quad (10)$$

公式中各项含义同公式(3)~(7)。为了使每个个体的适应度函数 $f(x)$ 的值都为非负,将适应度函数重新定义为:

$$f(x) = \frac{y}{1+g(x)} \quad (11)$$

遗传算法通过“优胜劣汰”选择个体,通过遗传算子操作可产生新的优良结构的个体,同时也存在把具有优良结构的个体破坏的能力,采用与适应值成比例的选择策略,既能使适应值高的个体具有更多的生存机会,但也因此又可能导致算法过早的不成熟收敛等,早熟产生的原因是进化的方向是向着高适应度的模式方向发展,造成模式缺少。如果能够进行有方向的基因补偿,可以解决基因丢失和基因浓度过低的问题(即早熟收敛)。

为了实现算法的自适应性并避免早熟现象,以自适应的惩罚函数^[6]的方式调整适应度函数;根据连续 k 代具有高适应度的个体满足约束条件的情况调整惩罚因子 λ 进而调整适应函数。

$$\lambda = \begin{cases} \lambda^* (1 - \delta), & \text{所有个体都满足条件时,} \\ \lambda^* (1 + \delta), & \text{没有一个个体全部满足约束条件时,} \\ \lambda & \text{其他} \end{cases}$$

如果这 k 代个体满足所有的约束条件,将 λ 乘以 $1 - \delta$ 。如果对于所有个体,没有一个个体全部满足约束条件,则将 λ 乘以 $1 + \delta$ 。其他情况则保持 λ 不变。以同样的方式调节惩罚因子 μ 。

3.2.3 控制遗传算法的参数和变量 一个简单的遗传算法由复制、杂交、变异三个遗传算子组成。控制遗传算法的主要参数有群体规模 N 和算法执行的最大代数 M ,次要参数有复制概率 p_r 、杂交概率 p_c 和变异概率 p_m 等参数。

种群由适应函数所对应的概率分布以轮盘赌形式确定^[7],即计算群体中个体的适应函数 $f(S)$,把当前群体中的个体按与适应值成比例的概率 $P_i(S)$ 复制到新的群体中,用于提高群体的平均适应值,其中

$$P_i(S) = \frac{f_i(S)}{\sum_{S \in P} f(S)} \quad (12)$$

$f(S)$ 是适应函数。

遗传杂交算子(有性重组)有多种,其中最简单的是一点杂交算子,即按概率 p_c 从群体中随机取出两个字符串,设串长为 L ,产生一个1到 $L-1$ 之间的随机数 i ,将两个串的右半段互换再重新连接得到两个新串。如:

$$\begin{cases} \text{父代 A } 10 \dots 01 \dots M \quad 01 \dots 10 \\ \text{交配位 } i \\ \text{父代 B } 01 \dots 00 \dots M \quad 10 \dots 11 \\ \text{子代 A } 10 \dots 01 \dots M \quad 10 \dots 11 \\ \text{交配位 } i \\ \text{子代 B } 01 \dots 00 \dots M \quad 01 \dots 10 \end{cases}$$

变异算子以一个很小的概率 p_m 随机地改变染色体串上的某些位,对于二进制串,就是相应的位从1变为0或从0变为1。如:

$$\begin{array}{c} 100 \dots 01 \dots 11 \dots 0 \\ \downarrow \\ 100 \dots 00 \dots 11 \dots 0 \end{array}$$

3.2.4 停止标准 本系统采用下面三种类型的停止标准:

最简单的标准是在固定循环次数后停止。这个固定次数依赖于模型复杂度,由题库中题目的个数 I 与模型中约束条件的个数 K 共同决定。本系统中 IK 次循环后停止;

另一种停止标准是基于群体中收敛的程度。即已生成的所有个体的平均适应度与最好个体的平均适应度的比大于等于0.995则GA停止。

第三种停止标准是算法执行一段时间后,如果没有新的更好解产生,则停止。

4 实验模拟与分析

本文所采用的模拟题库是基于三参数模型($a_i \sim U(0.5, 1.5)$; $b_i \sim N(0, 1)$; $c_i = 0.2$)的英语题库,具有450道题目。题库根据内容被分为三个子库。题目1~150是词汇题目;题目151~300是语法题目;题目301~450是阅读理解题目。每个子库中的前80道题目都是选择题,其他的是匹配题。测验要求在能力水平为 $\theta = -1$; $\theta = 0$; $\theta = 1$ 上的相关信息量为1($r_q = 1, q = 1, 2, 3$),测验包含10个词汇题、10个语法题、10个阅读理解题;15个选择题,15个匹配题。将组卷问题按照最大化模型(3)描述为:

maximize y ,

s.t.

$$A_1 \begin{bmatrix} x \\ y \end{bmatrix} \geq 0, \quad (13)$$

$$A_2 x = b_2, \quad (14)$$

$$x_i \in \{0, 1\}; i = 1, 2, \dots, I$$

$$y \geq 0,$$

其中

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_1 \\ M \\ x_I \\ y \end{bmatrix},$$

$$A_1 = \begin{bmatrix} I_1(\theta) & \dots & \dots & I_1(\theta) & -r_1 \\ M & & & M & \\ I_1(\theta) & \dots & \dots & I_1(\theta) & -r_2 \end{bmatrix}$$

150150150

$$A_2 = \begin{bmatrix} 1 & \dots & 10 & \dots & 00 & \dots & 0 \\ 0 & \dots & 01 & \dots & 10 & \dots & 0 \\ 0 & \dots & 00 & \dots & 01 & \dots & 1 \\ 1 & 10 & 01 & 10 & 01 & 10 & 0 \\ 0 & 01 & 10 & 01 & 10 & 01 & 1 \end{bmatrix},$$

$$b_2 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 15 \\ 15 \end{bmatrix},$$

例子中有五个约束条件,即 $K=5$, 计算适应度函数: $f(x) = \frac{y}{1+g(x)}$, 其中

$$g(x) = \lambda \sum_{i=1}^K h(A_2(i) * x - b_i),$$
$$h(u) = \begin{cases} u, & u > 0, \\ 0, & u \leq 0 \end{cases}$$

对上述描述的组卷问题应用自适应遗传算法,计算群体中每个个体的适应函数,对具有高适应度的个体进行复制、杂

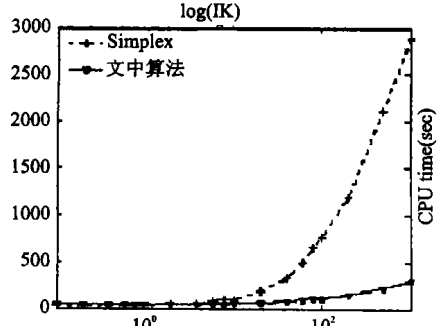


图 3 CPU 计算时间比较

Fig.3 Comparison of CPU time (in secs)

交、变异操作,同时检测连续 k 代具有高适应度的个体满足约束条件的情况调整惩罚因子 λ 如图 3 所示,比较了本文算法

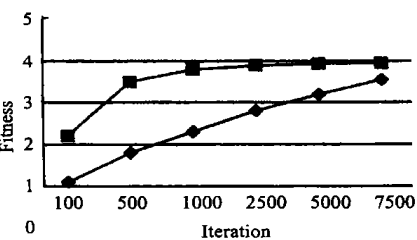


图 4 Fitness 比较

Fig.4 Comparison of Fitness

与解决线性规划与优化问题的传统方法——单纯形法 (Simplex)^[8]所需的 CPU 计算时间.我们看到当题库中的题

目数 I 越大,约束条件数 K 越多时,文中算法明显比 Simplex 算法耗费少得多的 CPU 时间.

自适应遗传算法应用惩罚因子调整适应函数,如图 4 所示,与简单的遗传算法相比,它能较早地找到满足条件的群体,最大程度地满足组卷要求,有效地解决了 IRT 组卷问题,为约束优化问题解决提供了一个新的有效的途径.

5 结 论

智能组卷问题是典型的约束组合优化问题,目前国内采用的是基于 CTT 随机抽取试题组卷或回溯法生成试卷的方式,国外采用的是应用线性规划法^[8]进行基于 IRT 组卷.本文研究了一个基于项目反应理论 IRT 的智能组卷方法——自适应遗传算法,该算法通过调整惩罚因子自适应地调整适应函数,有效地解决了组卷中的约束优化问题.基于 IRT 的智能题库系统是教育测量发展的趋势,国内在这方面的研究还很薄弱.

References:

1 http://www.uts.psu.edu/Classical_theory_frame.htm.

2 Qi Shu-qing, Dai Ha-qing, Ding Shu-liang. Modern Education and psychometrics [M]. Nanchang: JiangXi Education Press, 1998, 12.

3 Huub Verstralen, Timo Bechger & Gunter Maris. The combined use of classical test theory and item response theory [EB/OL]. http://www.cito.nl/pok/poc/eind_fr.htm, September 14, 2001.

4 Wim J. van der Linden, Ellen Boekkooi-Timminga. A maximin model for test design with practical constraints[J]. The Psychometric Society, June 1988, 53(2): 237~247.

5 De Jong K and Spears W. Using genetic algorithms to solve NP-complete problems [C]. Proceedings of the Third International Conference on Genetic Algorithms, 1989.

6 David W. Coit & Alice E. Smith, David M. Tate. Adaptive penalty methods for genetic optimization of constrained combinatorial problems[J]. Informs Journal on Computing, Spring 1996, 8(2): 173~182.

7 Shi Zhong-zhi. Advanced artificial intelligence [M]. Beijing: Science Press, 1998, 6.

8 Wim J. van der Linden & Bernard P. & Veldkam P. An integer programming approach to item bank design[J]. Applied Psychological Measurement, June 2000, 24(2): 139~150.

附中文参考文献:

2 漆书青,戴海崎,丁树良.现代教育与心理测量学[M].南昌:江西教育出版社,1998,12.

7 史忠植.高级人工智能[M].北京:科学出版社,1998,6.