

基于多策略的个性化智能组卷的研究

孙蓓蕾, 陈高云

(成都信息工程大学软件工程学院, 四川 成都 610225)

摘要:分析传统智能组卷所包括的各项要求以及各项约束条件,在用户个性化的基础上分析用户行为,将构造试卷的各项参数与用户的各项实际要求结合起来,使用人工鱼群算法进行智能组卷,充分凸显试卷的个性化特点,构建一种基于多策略的个性化智能组卷的理论模型,为提升教育教学效率及水平而成立,其专注纠错,深度挖掘用户的需求,基于错题库的数据分析,开展智能化地组卷训练,高精度度、高效率地解决用户查漏补缺的问题,并通过一系列的实例分析,通过实验结果表明提出的个性化组卷方法能够促使用户掌握自己的薄弱环节,在成绩方面有显著提高。

关键词: 计算机应用;智能组卷;人工鱼群算法;个性化;理论模型

中图分类号: TP315

文献标志码: A

0 引言

随着互联网的普及,网络教学已经在教学过程的各个阶段发挥日益重要的作用,而在网络教学中,在线学习与考试是非常重要的一部分,由此智能组卷成为该领域的一个热点^[1]。智能组卷是凭据出卷人给定的约束条件(目前考虑试题总数目、总分、知识点比例、难度系数、题型比例等各项因素),寻找题库中与符合各项约束条件的题目,组成最优的最满足需求的试卷。在实际组卷中,要生成一份符合大纲要求、难易适宜、题型比例合理的试卷是一个不断优化选择的过程,由于不同用户的基础和薄弱环节并不相同,通用的智能组卷并不能满足个性化这一需求^[2]。论文在智能组卷的过程中加入个性化因素,能针对性的对各项参数的分配系数进行合理量化,从而形成一套个性化的试卷。

1 常见的智能组卷策略

1.1 随机抽取法

随机抽取法是利用数学模型,将组卷的各项要求进行定义然后由计算机选择题目的算法过程,再从已有的试题库中随机抽取试题,符合用户要求的加入到试卷,不符合的则跳到下一道题,不断重复此过程直到生成一套试卷为止。此算法的优点是过程较为简单,易实现,早期的智能组卷较多使用这种算法。例如文献^[3]使用此方法,用集合的方式描述试卷的各项参

数,并且满足试卷要求的试题也通过集合来组合。

1.2 回溯试探法

回溯试探法是在随机抽取法的基础上进行改进,采用回溯的方法废除最近的一次操作,从某个回溯点重新往下搜索进行组卷^[4]。与随机算法比较,回溯算法在一定程度上减少抽取试题盲目性,缺点则是如果越是指标比较多的情况下将会消耗较多的时间,使得组卷的效率降低。

1.3 基于遗传算法的组卷

遗传算法是目前应用最广泛,且也是执行最有效的一种多约束目标优化算法,凭借简单通用,容易搜索到最优解等特点满足用户的智能组卷的需求^[5-6]。

2 基于多策略的个性化智能组卷模型

2.1 个性化智能组卷的步骤

在个性化组卷中,为了生成一套符合大纲要求并且满足用户个人学习特点的试卷,按照各个参数的权重比例,组卷的流程图如图1所示。

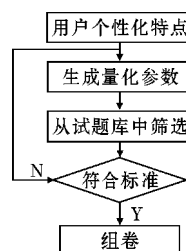


图1 流程图

2.2 组卷的约束条件

个性化试卷主要基于用户在一定行为基础上留下的错题库,在积累一定数量的错题后,展开错题分析和错题预测,并且跟考核大纲相结合,在纠错的同时保证试卷不偏离重点。因此组卷的约束条件主要有:试卷总分、知识点约束、题目类型约束、试卷难易度、易错题比例,文中通过五维空间对试题库进行约束实现个性化智能组卷,那么可以确定试卷的目标状态矩阵为

$$P=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&\vdots&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix}$$

其中:其中: m 为试卷中全部试题的总数; n 为每道考题的属性个数。目标矩阵 P 中行代表题目的属性,列则代表试卷中所有题目的所有属性。最终试卷时约束条件至多为为 $n-1$ 个,其中 $a_{il}(i=1,2,\cdots,m)$ 为试题的 ID 编号。
当所选择的题目符合了这些约束条件才能够被加入到试卷中,所有的约束条件如表 1 所示。

表 1 约束条件表

约束条件名称	公式	说明
总分约束	$\sum_{i=1}^m a_{i2} = S_p$	每道题目的分数属性 a_{i2} 的加和即为 S_p 是试卷的总分。
知识点约束	$\sum_{i=1}^m a_{i3}k(j) = S_k^j$	第 j 个知识点的分值约束为每一道归属于第 j 个知识点试题的分数 a_{i3} 的加和。其中: $k(j) = \begin{cases} 0, j \neq a_{i3} \\ 1, j = a_{i3} \end{cases}$, S_k^j 是第 j 个知识点的总分值。
题型约束	$\sum_{i=1}^m a_{i4}t(j) = S_t^j$	第 j 个题型的分值约束为每一道归属于第 j 个题型试题的分数 a_{i4} 的加和。其中: $t(j) = \begin{cases} 0, j \neq a_{i4} \\ 1, j = a_{i4} \end{cases}$, S_t^j 是第 j 个题型的总分值。
难度约束	$(\sum_{i=1}^m a_{i2}a_{i5})/S_p = DIF_p$	每道试题的分值与该试题的成绩的加和比上总分 S_p 等于难度约束值 DIF_p ,其中 DIF_p 根据用户留下的练习的平均成绩决定。
易错题约束	$\sum_{i=1}^m a_{i6}y(j) = S_y$	所有易错题的分值 a_{i6} 加和即为易错题的分值约束,其中: $y(j) = \begin{cases} 0, \text{非易错题} \\ 1, \text{易错题} \end{cases}$, S_y 是易错题的总分值。

2.3 个性化智能组卷的人工鱼群算法设计

人工鱼群算法就是根据鱼类的各种行为活动,通过构造人工鱼来模仿鱼群的觅食、聚群及追尾行为,从而实现寻优,最后使全局最优值在群体中得到显现,在多约束多策略求解问题中,鱼群算法能取得满意的效果^[7]。

人工鱼群算法对于较低维的问题,可以较好的跳出小范围极值、快速搜索到整个范围内的最优值,并且该算法有很多的优点,比如对初始值和参数的选取要求不高,算法稳定性比较高,而且算法更新机制简单、易操作也方便改进以应用于实际问题^[8]。

算法的基本思想为:目标函数的可行域即为一水域,而水域中的每条个体鱼即为目标函数的一个可行解,模拟鱼群觅食等行为,即可实现全局寻优^[9],结合智能组卷每一条人工鱼则代表一种试题组合方案,通过一次次的寻优过程,最后得到试题的最优组合方式也就是最符合约束条件的试卷。

2.3.1 预处理

在使用人工鱼群算法组卷前,需要将试题库的数据进行预处理工作,首先将试题按照题型进行分类,同

一题型的试题组建为一个类别,由此将试题库分成多个不同的类别,不同的题型之间进行编号处理^[10]。

2.3.2 编码方式

每条鱼的编码方式采用二进制编码方法,使用 m 维的向量 $X=(X_1,X_2,\cdots,X_m)$ 表示。由试卷的数目决定向量的所需维数,向量的分量 $x_i(i=1,2,\cdots,n)$ 表示试题属性, n 为试题属性的个数。

2.3.3 标函数

目标函数是人工鱼群算法中用来计算食物浓度,在符合各项条件的试卷中的总分、知识点比例、题型比例、试卷难度和易错题比例这些约束条件在理想状态下应该与设定值完全一致或者在一定的误差范围内可以被接受^[11]。生成试卷的各项参数和设定值之间进行加权计算的结果决定了目标函数 Y 值的大小。只有当 Y 值最小的时候,生成的试卷才是符合各项要求的试卷,也就是食物浓度高的状态。目标函数按照下面公式进行计算

$$Y=\left|\sum_{j=1}^k\left(\sum_{i=1}^ma_{i3}k(j)-S_k^j\right)\right|\cdot w_1+\left|\sum_{j=1}^t\left(\sum_{i=1}^ma_{i4}t(j)-S_t^j\right)\right|\cdot w_2+$$

$$\left| \left(\sum_{i=1}^m a_{i2} a_{i5} \right) / S_p - DIF_p \right| \cdot w_3 +$$
$$\left| \sum_{j=1}^y \left(\sum_{i=1}^m a_{i6} y(j) - S_y \right) \right| \cdot w_4$$

没有对试卷总分进行约束,是由于在初始化过程中将总分设置为定值,公式中 k 代表课程的总的知识点数目, t 表示总的题型属性数目。

2.3.4 算法步骤

步骤 1:人工鱼参数初始化。假定人工鱼的种群数目为 m ,每条人工鱼的维数为 n ,对每条人工鱼的位置信息进行初始化操作。题库中的试题全部进行编号,比如按照题型分类,然后相同的类别进行规整,放入数组中进行管理,数组的连续性可以对题目进行编号操作。

步骤 2:确定试卷中每个题型的数目。每种题型都有一定的分数约束,通过这个分数约束自然可以计算出每种题型所需要的题目数目,通过这个约束条件就可以使得初始化过程中保证每个题型的总和分数满足约束需求。

步骤 3:设置参数。对鱼群的状态进行初始化,首先从题库中进行抽取试题的操作,题目的数量符合约束条件的要求,同时设置鱼群之间的最大允许距离为 V ,人工鱼的最大移动步数为 S ,最大的拥挤因子为 δ 等参数。而鱼群中的食物浓度通过目标约数函数 Y 进行表示, Y 值越小,说明人工鱼所处的状态越高,生成的试卷越符合各项约束条件。

步骤 4:初始化公告板。通过当前鱼的状态所对应的试题编号,计算相关属性是否符合目标函数约束值,将当前状态下目标函数值最小的状态值以及对应的目标函数值都在公告板上进行记录和对旧值更换,并且等待与下一条鱼的值进行比较。

步骤 5:更新公告板。鱼群中的人工鱼都要进行对比操作,由此筛选出状态最佳也就是目标函数值最小的人工鱼的状态,例如进行追尾行为模拟和聚群行为模拟,以此更新迭代,从而使公告板上的值为最优值,如果小于则进行更换,反之则不做改变。

步骤 6:终止。重复步骤 4 和步骤 5,当公告板上的目标函数值为最小为止,或者是比较的循环次数大于了约束条件,然后输出最优的试卷。

比如针对某用户进行个性化组卷,设置总分约束为 100 分,各个知识点和题型的约束情况见表 2、表 3。

表 2 知识点的约束情况

知识点	一	二	三	四	五	六
比重	5	15	40	20	10	10

表 3 题型的约束情况

题型	分值	题数
单选	2	10
多选	4	5
填空	2	10
简答	8	5

分析用户的薄弱环节,获取他的错题库,进行分类归纳计算出错题约束的各项比重,带入到上述算法中,进行迭代循环指导找到最符合各项条件的试卷。

3 实验结果与分析

将多策略的个性化智能组卷模型与基于遗传算法的组卷模型^[12]的试卷的知识点分布进行对比分析,前者生成的试卷的题目对应的知识点相对集中,后者则较为分散,如图 2 所示,图 2(a)、(b)是针对学生的个性化特点组成的试卷,知识点分布较为聚集,图 2(c)为传统组卷方式,不同的学生组出来的试卷知识点分布情况基本相同。

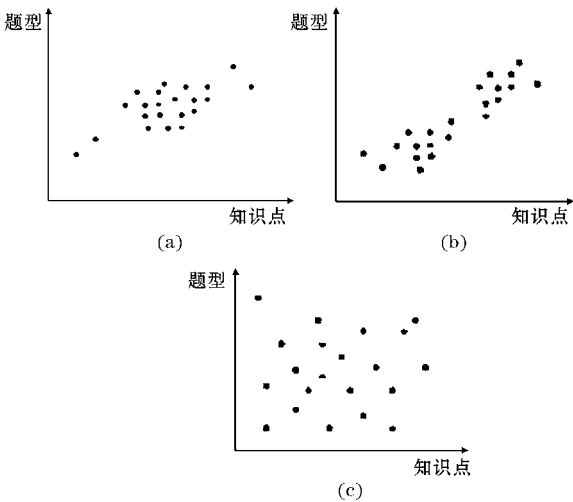


图 2 分布图

将基于多策略的个性化智能组卷模型应用于课程教学过程化管理平台,在教学班级中选取一部分学生使用平台原来使用一般方法组成的试卷进行练习,一部分学生使用针对自身薄弱环节的个性化试卷进行练习,经过一段时间的测试,两部分学生最开始的成绩平均分均为 55 分,经过一定时间的测试后,平均成绩增长速度并不一样,实验结果如图 3 所示。

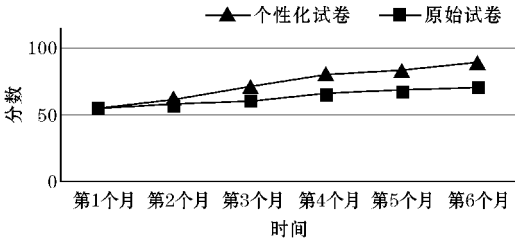


图 3 平均分增长结果表

由实验结果可知,以个性化为出发点的智能组卷从用户的错题集着手,对已有的试题库进行初始化处理,提高筛选试卷题目的速度;相对使用平台以往的普通的试卷的用户来说,在相同的时间内,使用符合个人特点的个性化试卷的用户提高成绩的速度较快和涨幅更高,并且随着时间的推移,优越性更强。

4 结束语

针对试卷使用者的个性化特点,对个性化试卷生成问题进行设计与算法研究,提出基于多策略的个性化智能组卷的模型。方法组成的试卷不仅覆盖各个科目的大纲需求,并且在试卷难度上也是根据用户的平时成绩进行循序渐进,能够促使用户进行查漏补缺,更好地提高学习水平和学习效率。

参考文献:

- [1] 冯素梅. 通用试题库管理系统的分析与设计[J]. 齐齐哈尔大学学报:自然科学版, 2013, (1):26-30.
- [2] 唐艺, 梁刚. 基于组合优化的高校体育理论题库的研究与设计[J]. 湖南师范大学学报:自然科学版, 2011, (5):46-49.
- [3] 潘志国. 分段随机抽选算法在工程图学智能试题库系统设计中的应用[J]. 电子测试, 2015, (8):145-147.
- [4] 范晶晶, 冯英伟, 周丽莉. 回溯试探组卷算法及网络问卷质量保证研究[J]. 河北建筑工程学院学报, 2015, (2):55-57.
- [5] YANG Jian, ZHANG Min-hui. Research on an Intelligent Test Construction Method Based on Improved Genetic Algorithm[J]. Computing Technology and Automation, 2012, (2).
- [6] 唐启涛. 基于改进的遗传算法的智能组卷算法研究[J]. 计算机技术与发展, 2014, (12):241-244.
- [7] 李娟, 田胜利. 基于人工鱼群模型的自动组卷抽题算法研究[J]. 许昌学院学报, 2008, (5):92-97.
- [8] 任剑, 卞灿, 全惠云. 基于层次分析方法与人工鱼群算法的智能组卷[J]. 计算机应用研究, 2010, (4):1293-1296.
- [9] Cai Yun. Artificial Fish School Algorithm Applied in a Combinatorial Optimization Problem[J]. International Journal of Intelligent Systems and Applications 2010, (2):37-43.
- [10] 俞洋, 殷志锋, 田亚菲. 基于自适应人工鱼群算法的多用户检测器[J]. 电子与信息学报, 2007, (1):121-124.
- [11] 窦昱, 赵洁, 黄静. 自测系统中基于多策略的组卷方法研究[J]. 东北农业大学学报, 2007, (4):539-541.
- [12] 鲁萍, 何宏璧, 王玉英. 智能组卷中分级带权重知识点选取策略[J]. 计算机应用与软件, 2014, (3):67-69.

Studies on the Intelligent Composing Paper with Multi-strategy and Individuality

SUN Bei-lei, CHEN Gao-yun

(College of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Through the analysis of the requirements and constraint conditions of the traditional intelligent test paper composition and the user behavior on the basis of individuation, a theoretical model based on the multi-strategic personalized intelligent test paper composition was established, which combines the parameters in the paper with the requirements of users and generate the paper with the help of AFSA to fully prominent its characteristic of individuation. This theoretical model aims to enhance the teaching efficiency, which focuses on error correction and deeply explore the users' demand. Based on the analysis of the error data bank, the intelligent test paper composition can help the users solve their error accurately and effectively. Finally, case analysis shows that this method of generating test paper is able to prompt users to overcome their weakness.

Key words: technology of computer application; intelligent test paper composition; artificial fish school algorithm; individuation; theoretical model.