

基于智能算法的自动组卷在线考试系统

张 健

(宿迁泽达职业技术学院, 江苏 宿迁 223800)

摘 要: 目前, 很多在线考试系统已基本实现了在线考试、自动组卷和成绩统计等功能, 但传统的自动组卷算法存在效率低、随机性高、试卷难易不易控制等问题。针对以上问题, 笔者分析自动组卷的相关理论, 设计了一款功能全面、性能优越的基于智能算法的自动组卷在线考试系统。

关键词: 在线考试系统; 自动组卷; 智能算法

中图分类号: TP183 文献标识码: A 文章编号: 1003-9767 (2021) 07-097-03

Online Examination System of Automatic Paper Composition Based on Hybrid Algorithm

ZHANG Jian

(Zeda Vocational and Technical College of Suqian, Suqian Jiangsu 223800, China)

Abstract: At present, many online examination systems have basically realized the functions of online examination, automatic paper composition and score statistics, but the traditional automatic paper composition algorithm has some problems such as low efficiency, high randomness, and difficult test paper control. In view of the above problems, this paper analyzed the related theories of automatic paper composition, and designed a comprehensive and superior performance based on hybrid algorithm automatic paper composition online examination system.

Keywords: Online examination system; Automatic paper composition; Hybrid algorithm

0 引言

21 世纪以来, 计算机软硬件技术、信息技术处理、大数据处理、人工智能等进入高速发展的阶段, 自动组卷在各个领域得到迅猛发展。新冠疫情期间, 在线考试成为重要的考试形式, 其中卷算法是合成试卷的成功率和有效率的关键。目前, 主要的自动组卷算法有随机抽取算法、误差补偿算法、回溯试探法和遗传算法等^[1]。本文主要介绍基于改进遗传算法的自动组卷在线考试系统, 将试卷内容的约束条件映射到遗传算法的表达空间, 同时采用子群的精英保留策略, 大大加快了组卷过程中的收敛速度。

1 自动组卷的概念

自动组卷指根据限定的试题数目、题型、分数、知识点范围、考点分布等因素, 自适应地搜索题库, 根据每个限定因素的特征参数的权重匹配出最优试卷, 主要用于解决多重约束的组合优化问题^[2-3]。本文提出的基于混合算法的在线考试系统组卷参数有考题总分、每类题型包含的考题数目、各

个题型对应的分数、各个题型的难度系数以及各个章节占据的分数。组卷的时候参照以上 5 个制约要求, 从题库内抽取一道考题, 基于整体范畴的自适应遗传算法进行自动组卷, 根据数据库表关系对题型分段式编码, 在算法迭代过程中使用交叉、变异算子动态自适应和子群的精英保留策略, 大大加快了遗传算法的收敛速度, 提升了组卷效率和试卷质量。

2 改进遗传算法在自动组卷中的应用

2.1 算法概述

遗传算法最早于 20 世纪 70 年代提出, 算法的核心是将实际问题转化成数学函数最优解的问题^[4]。算法实现的过程是以编码的形式来处理二进制的染色体, 模拟二进制进行种群的多次交叉和变异等过程, 然后生成问题的最优解。这一过程产生的每一代都将继承父代的优良基因, 即保留适应度函数较高的解, 淘汰适应度较低的解, 以生成子代的二进制字符串^[5]。通过模拟生物的遗传进化, 遗传算法被证明是各种仿生系统中适应优化能力最强的算法^[6]。

作者简介: 张健 (1988—), 男, 江苏宿迁人, 本科, 讲师。研究方向: 计算机应用、自动化。

2.2 编码方式与初始种群

遗传算法在解决不同的组卷问题时,编码方案决定算法的性能。编码方案是将问题空间转化到遗传算法空间。主流的方式包括二进制、十进制和自定义的实数编码方案,但随着试题量的增大会导致编码冗长、占用内存等情况^[7]。基于上述情况,本文根据数据库表结构采用分段式编码方案^[8],这套编码方案是采用独立分段编码的方式确保编码间互不干扰,在缩短组卷时间的同时,能够提高组卷的成功率。独立分段编码方案如表1所示。根据题型以及数据库表关系独立实数分段编码,通过设定题型在试卷中的比例,达到缩短组卷时间以及提高组卷成功率的目的。

表1 独立分段编码方案

题型	分段编码
单选题	26、4、34
判断题	81、2
简答题	66

2.3 适应度函数

遗传算法中的适应度函数是决定组卷性能好坏的关键,组卷过程中的适应度函数值越大越可以指导问题搜索的进化方向。适应函数有两个基本要求:变量输入的非负性以及目标函数对试题搜索起正反馈。试卷的不同组成部分对应于试卷的不同属性,各个属性在试卷中的权重也不同,记属性*i*的权重为 t_i ,得到属性和如式(1)所示:

$$\sum_{i=1}^n t_i R_i \quad (1)$$

对应于适应度函数,如式(2)所示:

$$f(x) = \frac{1}{1 + \sum_{i=1}^n t_i R_i} \quad (2)$$

式中, R_i 是组卷过程中各个属性的期望值与实际值的偏差,表示为 $R_i = |E_i - P_i|$, R_i 值越大,适应度函数值也越小,代表选中的试题越不符合目标要求。

2.4 遗传算子改进

2.4.1 选择算子

选择算子目标在于淘汰不符合要求的个体和保留最优个体,通过多次迭代达到去除劣质个体,在实际组卷过程中会根据适应度函数的大小对试题进行筛选。常用的选择算子包括轮盘赌选择和锦标赛选择等方式,其中轮盘赌选择会保留部分劣质个体以保证群体的多样性以及避免陷入局部最优,但会遇到无法收敛的问题^[9]。锦标赛选择则只保留最优个体,无法保证群体的多样性和全局最优解。因此本文采用轮盘赌选择和锦标赛选择相结合的方式, n 个群体的总适应度可表示为:

$$S = \sum_{i=0}^n f(x_i) \quad (3)$$

个体被选中的概率可表示为:

$$P(x_i) = f(x_i) / S \quad (4)$$

2.4.2 交叉算子和变异算子

交叉算子采用的是分段式单点交叉的策略,群内个体根据交叉概率 P_c 和满足条件的交换随机两两配对组合,以遗传的方式生成新的个体进入下一次迭代过程。本文经过验证发现, P_c 的取值为0.30~1.0时的效果最佳。

变异算子主要为了维持种群的多样性,当产生新的个体后会有很低的变异概率 P_m ,主要操作体现在使用rand(0,1)随机函数生成数判断与变异概率比较,如果生成的随机数小于变异概率,则个体产生变异,否则保持原状。

交叉概率 P_c 和变异概率 P_m 对新个体产生的影响巨大,直观表现在交叉概率 P_c 越大,个体间交叉组合概率越大,产生的新个体也越多,会导致难于收敛的情况。交叉概率 P_c 越小就会导致搜索时间变长;变异概率 P_m 越大,种群多样性越大,也会导致退化成随机抽取算法,变异概率 P_m 越小,会导致数据欠拟合和局部最优^[10]。本文采用交叉概率 P_c 和变异概率 P_m 自适应取动态最优的方案。

2.5 最优个体保留与终止策略

在遗传算法每一轮的迭代以后,在一定概率下会淘汰种群中的最优个体。因此本文采用精英保留的策略保留上次种群迭代的最优解,不参与下一次的迭代交叉变异过程。使用这种策略,在保留群体总适应度的同时也加快了遗传算法的收敛速度。遗传算法在执行过程中,由于初始种群的差异性以及编码方式不同导致无法判断种群迭代是否达到最优解,所以设置最优的个体终止策略尤为重要。

系统执行过程是多次迭代和对空间解不断优化过程。组卷系统终止条件主要考虑以下3个方面:①群体适应度总和达到预设阈值的某个区间范围内时;②循环次数达到最大迭代次数时;③连续几代种群平均适应度达到预设阈值的某个区间范围内时,算法终止。

2.6 算法的执行过程

①获取题库:首先获取题库的试题信息,从题库中抽取试题。②初始化试卷:初始化组卷需要的基本条件,比如试卷的科目、总分、题量、题型、难度要求和考点信息等。③初始化数据:产生初始种群,设置算法的最大迭代次数、保留群体个体数、适应度的期望值、交叉概率和变异概率等。④计算适应度值:开始迭代,根据适应度计算公式计算个体适应度值。⑤判断适应度值和最大迭代次数:如果得到的试卷个体的适应度值达到期望值则停止迭代和生成试卷;如果未达到期望适应度值,根据精英保留策略,将适应度值最高的个体保留到子代种群中;同时判断程序迭代次数是否超过设定的最大迭代次数,如果超过则结束程序,组卷失败。⑥改进遗传算子操作和生成新的种群:对种群个体进行选择算子、

交叉算子和变异算子操作,生成新的种群。⑦继续执行:跳到步骤③计算个体适应度,继续运行程序。具体的组卷流程如图1所示。

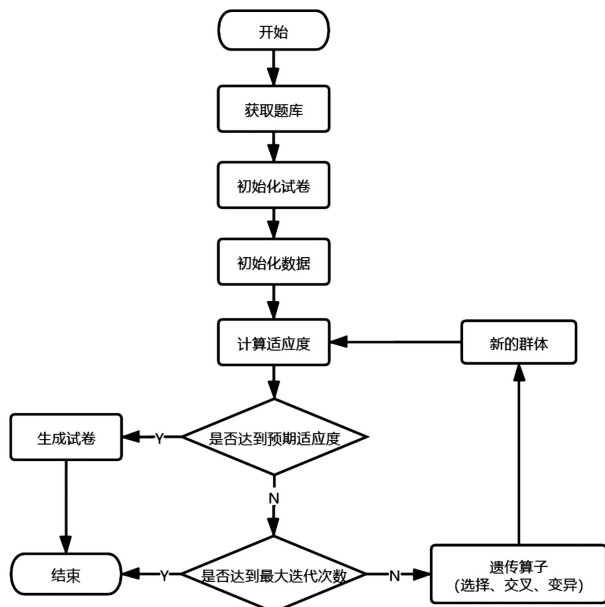


图1 改进遗传算法的组卷流程

3 实验与分析

通过系统后台对题库进行组卷模拟,初始化试卷信息参数:题型3个,包括3道填空题、2道判断题和1道简答题,设置填空题和判断题每道6分,简答每道20分,总分50分,设置难度信息为0.5(中等)。设置改进遗传算法的参数:保留每一代个体数3个,设置交叉概率 P_c 为0.55和变异概率 P_m 为0.02,程序最大迭代次数300,期望适应度值0.99。采用改进后的基于智能算法的自动组卷与传统组卷算法,得到3对组卷效果,如表2所示。从表2可以看出,与传统遗传算法的组卷效果相比,改进遗传算法的自动组卷算法的程序迭代次数和执行时间明显减少,试卷的收敛速度和组卷效率得到提高,进一步突出了本方法的优越性和适用性。

表2 遗传算法与传统算法效果对比

试验名称	传统遗传组卷算法		改进遗传算法的组卷算法	
	程序迭代次数	组卷时间/s	程序迭代次数	组卷时间/s
1	211	65	137	43
2	199	58	101	38
3	187	56	81	29

4 结语

在线考试是教学过程中的一个重要环节,也是重构教学组织与策略的重要环节之一,而在线考试中,自动组卷算法是其中的核心。本文从自动组卷的概念出发,通过介绍遗传算法的基本原理对遗传算子进行改进,并将其应用在自动组卷过程中,取得较好的效果。

参考文献

- [1] 张亚妮. 对传统几种组卷算法改进与比较分析[J]. 计算机与数字工程, 2017, 45(12): 2364-2367.
- [2] 舒清录. 一种实用的随机组卷策略研究与实现[J]. 电脑与信息技术, 2020, 28(3): 30-32.
- [3] 孙俊丽. 试题库系统的组卷算法分析与研究[J]. 办公自动化, 2019, 24(13): 48-49.
- [4] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 1999: 30-32.
- [5] 杜洋. 遗传算法原理和应用[J]. 才智, 2010(9): 53.
- [6] 黄超叶, 卢光云, 覃仲欣. 智能组卷试题库系统的设计与开发探讨[J]. 信息通信, 2020(9): 76-78.
- [7] 张三群. 计算机应用基础考评系统(以Word模块为例)的设计与实现[J]. 信息技术与信息化, 2020(1): 118-120.
- [8] 韩啸. 基因表达式编程算法的改进及其应用[D]. 大庆: 东北石油大学, 2020: 15.
- [9] 刘雅莉. 一种改进遗传算法的自动组卷系统优化研究[J]. 微型电脑应用, 2020, 36(8): 28-30.
- [10] 沈林玉, 叶恒舟. 基于遗传算法的自测系统设计与实现[J]. 电脑与信息技术, 2021, 29(1): 1-3.