

分类号\_\_\_\_\_

密 级\_\_\_\_\_

UDC\_\_\_\_\_

论文编号\_\_\_\_\_

# 硕士学位论文

( 专 业 学 位 类 型 )

论文题目：遗传算法在学科平台系统专家  
分配中的应用

研 究 生：\_\_\_\_\_

导 师：\_\_\_\_\_

研究方向（领域）：行业信息系统

2019 年 3 月

分类号：

学校代号： \*\*

学 号：

秘密☆ 年

\*\*大学硕士学位论文

遗传算法在学科平台系统专家分配中的应用

作者姓名：

指导教师姓名、职称：

申请学位类别： 工程硕士

学科专业名称： 计算机技术

研究方向： 行业信息系统

论文提交日期： 年 月 日

论文答辩日期： 年 月 日

学位授予单位： \*\*大学

学位授予日期： 年 月 日

答辩委员会主席： \_\_\_\_\_

# **Application of Genetic Algorithm in Expert Distribution of Discipline Platform System**

A Thesis Submitted for the Degree of Master

**Candidate:**

**Supervisor:**

\*\* University

Wuhan, China



## 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立撰写而成。学位论文的核心内容为本人在导师指导下独立进行研究所取得的成果，文中引用的其他个人或集体的成果均以规范的方式注明。本人完全意识到本声明的法律后果由本人承担。

作者签名：

日期： 年 月 日

## 学位论文使用授权书

本论文作者完全了解学校关于保存、使用学位论文的管理办法及规定，即学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人完全同意《中国博士学位论文全文数据库出版章程》、《中国优秀硕士学位论文全文数据库出版章程》(以下简称“章程”，见 [www.cnki.net](http://www.cnki.net))，愿意将本人的学位论文提交中国学术期刊(光盘版)电子杂志社在《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》中全文发表和以电子、网络及其他数字媒体形式公开出版，并同意编入 CNKI《中国知识资源总库》，在《中国博硕士学位论文评价数据库》中使用和在互联网上传播，同意按“章程”规定享受相关权益(请作者直接与杂志社联系，电话：010—62791817、62793176、62701179；通讯地址：北京 清华大学邮局 84-48 信箱 采编中心 邮编：100084)。

学位论文作者签名：

导师签名：

年 月 日

年 月 日

注：本论文如需保密，保密级别是\_\_\_\_\_，解密时间是\_\_\_\_年\_\_\_\_月。(保密学位论文在解密后适用于本授权书)

# 遗传算法在学科平台系统专家分配中的应用

## 摘 要

同行评议的本质是评价主客观事物的一种方法，是当今科学界普遍认同的一种评价制度，已被广泛应用于论文发表、职称评议、学位点评议、成果评价、申请经费等科研评价活动中，推动着科研事业的蓬勃发展。专家遴选是同行评议的首要 and 基础工作，这就要求参与评议工作的专家团队必须合理、科学、公平，尽可能客观地反映实际情况，提高评议结果的质量。本论文研究的系统是学位与研究生教育管理信息系统（简称学科平台），是作者所在单位承接的一个实际工程项目，旨在建成一个简洁、安全、稳定、高效、智能的专家评议系统。

一般的专家分配是给每个项目分配相同数量的专家，这是一对多的专家分配问题，而学科平台系统需要的是为每一组学科分配相同数量的专家，即多对多的专家分配问题。当专家数量较少时，人工分配专家是不错的选择，然而随着专家数量的增加，分组数量较大时，那么人工分配专家和回溯算法就显得束手无策了，因此一个更加有效、科学、智能的专家分配方法是解决目前专家分配问题的重要途径。

首先，本文介绍了遗传算法的基本术语和操作步骤，并结合该系统中专家分配问题的特殊需求，提出了基于遗传算法的专家分配问题。

其次，根据学科平台系统分配专家的实际需求，建立了基于遗传算法的多目标优化的数学模型，逐步确立了遗传算法的编码方式、选择、交叉、变异、适应度函数等核心要素。

最后，实现了系统的核心功能，并将遗传算法成功地运用于专家遴选中。根据实验结果可知，结果优于传统的手动选择。实验结果表明该方法在实际应用中是有效的，从而证明了该模型的正确性和该方法的可行性。

**【关键词】：**专家分配；专家评议；数学模型；遗传算法

# **Application of Genetic Algorithm in Expert Distribution of Discipline Platform System**

## **Abstract**

The essence of peer review is a method of evaluating subjective and objective things. It is an evaluation system generally accepted by the scientific community today. It has been widely used in research publications, title evaluation, degree review, results evaluation, application funding and other research evaluation activities. In the middle, it promotes the vigorous development of scientific research. Expert selection is the primary and basic work of peer review. This requires that the team of experts involved in the review work must be reasonable, scientific, and fair, reflect the actual situation as objectively as possible, and improve the quality of the evaluation results. The system studied in this thesis is the degree and postgraduate education management information system (referred to as the disciplinary platform). It is an actual engineering project undertaken by the author's unit, aiming to build a simple, safe, stable, efficient and intelligent expert review system.

The general expert assignment is to assign the same number of experts to each project. This is a one-to-many expert assignment problem, and the disciplinary platform system needs to be for each group of disciplines (may be composed of the same discipline, or it may be Different disciplines are assigned to assign the same number of experts, that is, many-to-many experts. When the number of experts is small, manual assignment of experts is a good choice. However, as the number of experts increases and the number of groups is large, manual assignment experts and backtracking algorithms are helpless,

so a more effective, scientific, and intelligent the expert allocation method is an important way and inevitable method to solve the current expert allocation problem.

Firstly, the basic terms and operation steps of genetic algorithm are introduced. Combined with the special requirements of expert allocation problem in the system, the expert allocation problem based on genetic algorithm is proposed.

Secondly, according to the actual needs of the discipline platform system distribution experts, a mathematical model of multi-objective optimization based on genetic algorithm is established, and the core elements such as coding, selection, crossover, mutation and fitness function of genetic algorithm are gradually established.

Finally, not only the core functions of the system are realized, but also the genetic algorithm is successfully applied to the expert selection. According to the experimental results, the results are superior to the traditional manual selection. The experimental results show that the method is effective in practical applications, which proves the correctness of the model and the feasibility of the method.

**【Key words】** : expert assignment; expert review; mathematical model; genetic algorithm



# 目 录

第 1 章 绪论 .....	1
1.1 研究背景 .....	1
1.2 研究意义 .....	2
1.3 国内外研究现状 .....	3
1.4 本文研究内容及组织结构 .....	4
第 2 章 基于遗传算法的专家分配方法研究 .....	5
2.1 遗传算法基本原理 .....	5
2.1.1 遗传算法的基本术语 .....	6
2.1.2 基本遗传算法的操作步骤及流程 .....	7
2.2 学科平台系统中专家分配方法的研究 .....	9
2.3 基于遗传算法的专家分配方法的分析 .....	10
2.4 本章小结 .....	12
第 3 章 学科平台系统专家分配的需求分析 .....	13
3.1 学科平台系统的总体需求 .....	13
3.2 学科平台系统详细需求分析 .....	15
3.2.1 材料管理 .....	15
3.2.2 分组管理 .....	16
3.2.3 专家信息管理 .....	16
3.2.4 专家遴选 .....	17
3.2.5 专家评议 .....	18
3.3 学科平台系统设计 .....	19
3.3.1 与专家分配相关的类图设计 .....	19
3.3.2 与专家分配相关的表设计 .....	20
3.3 本章小结 .....	21
第 4 章 基于遗传算法的专家分配方法在学科平台系统中的应用 .....	22
4.1 专家分配问题的数学模型 .....	22

4.1.1 专家分配的问题描述 .....	23
4.1.2 材料和专家的处理 .....	24
4.1.3 分配专家的约束条件确定 .....	25
4.1.4 建立专家分配问题的目标函数 .....	27
4.2 遗传算法在专家分配方法中的应用 .....	28
4.2.1 确定编码方案 .....	28
4.2.2 群体规模与初始化 .....	29
4.2.3 适应度函数的设计 .....	29
4.2.4 遗传选择、交叉和变异 .....	30
4.2.5 遗传算法终止 .....	33
4.5 本章小结 .....	34
第 5 章 学科平台系统实现与遗传算法应用效果分析 .....	35
5.1 开发运行环境 .....	35
5.2 系统的实现 .....	35
5.2.1 材料管理的实现 .....	35
5.2.2 分组管理的实现 .....	39
5.2.3 专家信息管理的实现 .....	40
5.2.4 专家遴选的实现 .....	35
5.2.5 专家评议的实现 .....	40
5.3 应用效果分析 .....	40
5.3.1 实际效果测试结果 .....	41
5.3.2 基于遗传算法的专家分配方法适用性分析 .....	42
5.3 本章小结 .....	43
第 6 章 总结与展望 .....	44
6.1 论文完成的主要工作 .....	44
6.2 未来工作展望 .....	45
参考文献 .....	46
附录：攻读硕士学位期间发表论文及参与的项目 .....	50

# 图目录

图 2.1 遗传算法运算流程.....	7
图 2.2 基于遗传算法的专家分配方法流程.....	12
图 3.1 系统总体需求用例.....	13
图 3.2 材料管理功能.....	15
图 3.3 分组管理功能.....	16
图 3.4 专家管理功能.....	17
图 3.5 专家评议步骤.....	19
图 3.6 系统核心类类图.....	20
图 4.1 专家分配中的遗传算法流程图.....	22
图 5.1 基于遗传算法的专家分配页面.....	36
图 5.2 基于遗传算法的专家分配结果.....	36
图 5.3 遗传算法核心流程代码.....	37
图 5.4 遗传算法选择算子代码.....	37
图 5.5 遗传算法交叉算子代码.....	38
图 5.6 遗传算法变异算子代码.....	38
图 5.7 材料管理界面.....	39
图 5.8 分组管理页面.....	39
图 5.9 专家信息管理页面.....	40
图 5.10 专家评议界面.....	40
图 5.11 运行效果图.....	42

# 表目录

表 2.1 遗传算法应用领域及说明 ..... 5

表 2.2 常用分配专家方法对比 ..... 10

表 3.1 申请书材料表 ..... 20

表 3.2 专家表 ..... 21

表 3.3 专家分组及材料分组表 ..... 21

表 4.1 专家分配示例表 ..... 23

表 4.2 信誉等级量化处理表 ..... 24

表 4.3 专家与分组之间的学科紧密度量化处理表 ..... 25

表 4.4 专家与分组之间的总紧密度表 ..... 26

表 4.5 个体与个体适应度表 ..... 31

表 4.6 个体选择概率与累积概率表 ..... 32

表 4.7 个体编码表 ..... 32

表 4.8 均匀交叉算子示例表 ..... 33

表 4.9 单点变异算子示例表 ..... 33

表 5.1 开发环境所需工具表 ..... 35

表 5.2 实际权重参数设置表 ..... 41



## 第 1 章 绪论

### 1.1 研究背景

硕士、博士学位授权审核制度是学位制度的重要组成部分，目的是为了提高学位的质量，以及提高社会的信誉度，是一种学术自律的控制制度<sup>[1]</sup>。学位授权审核制度在 20 多年的发展过程中<sup>[2]</sup>，为我国教育评审积累了丰富的经验，同时也面临着新形势和新任务，因此需要进一步深化学位授权审核制度改革。国务院学位委员会印发了《博士硕士学位授权审核办法》<sup>[3]</sup>，决定 2017 年展开博士硕士学位授权审核工作，要求以质量作为标准，对于新增的学位授予单位和学位授权点，要制定起点更高、标准更严、针对性更强的申请基本条件。学位授权审核工作的目标是以服务社会发展的需求，必须保证学位授予的高质量，努力构建一个责权分明、统筹规划、分层实施、公正规范的制度体系<sup>[4]</sup>。根据不同的用户角色，学科平台系统需要提供不同的功能需求，参与评审工作的专家团队必须合理、科学、公平，能够结合不同学科点申报材料的特点和需求，提高专家组与学科点申报材料组的匹配度。

“同行评议”一词，由来已久，现指的是由某一或若干领域内的若干专家对上述领域的知识产品进行评价的活动，是一种评价事物的手段或方法，同时，也是一种群体决策的表现形式。随着当今科技的不断进步，尤其是计算机技术的出现<sup>[5]</sup>，使得人类进入了信息化革命时代，计算机可以代替人许多的复杂劳动，节省了大量的人力和物力，同时还提高了人们的工作效率。凭借着具有先天性优势的计算机技术，在各个领域发挥出了巨大的威力和价值，这就给同行评议带来了新的管理模式。

同行评议的第一步就是要为项目确定所需的评议专家<sup>[6]</sup>，专家的分配是项目评议的首要 and 基础工作。各省都有类似学科平台的系统，但是大多数的系统都落后于计算机技术的发展，仍然采用传统的人工分配专家的形式，特别是在专家分配方面，基本都是在海量的专家中人工选择专家，这种传统的分配专家方式，比较容易出现偏差，出现不合理、不公正、不科学的现象。如果分配的专家所熟悉的学科与分配的材料所属学科有很大出入，那么将会出现专家拒绝参与评议工作，或者出现专家评议质量不高的现象。目前，在专家拒绝参与评议工作时，解决的办法只有通过手动的方式替换专家，这样不仅耽误评议进度，同时造成不必要的人员浪费。因此传统的人工分配专家方式有巨大的进

步空间。于是，高效的分配专家方法已经成为省学位办关注的热点之一，改进分配专家方法，使得专家分配工作更加简单，满足博士硕士授权学位审核工作的实际需求，减轻学科平台系统工作人员的负担，提高评议的工作效率，保证专家分配过程的公平、合理、科学。

## 1.2 研究意义

随着“互联网+”在教育领域的不断深入，省学位办迫切地需要通过计算机减轻博士硕士授权学位审核相关工作人员的工作量，考虑到专家地理位置的限制，而且每位专家需要评议多份学科点申报材料，把专家集中在一个地方进行评议是浪费人力物力和财力的，因此，依托于计算技术的网上评议系统就很有必要。

初期，专家分配是为项目逐一分配相同数目的评议专家，分配的依据是专家熟悉的学科与项目的匹配程度，从提前准备好的专家库中一一遴选出对应学科的专家<sup>[7]</sup>。当分组比较少时，此方法完全满足用户的需求。

随着项目分组的增加，专家的数据量也越来越庞大起来，那么上述的方法就会出现相应的问题：

1. 分配专家的相关工作人员劳动量增大，且主观性较强，可操控的空间较大。
2. 如果分配专家数量过多，那么就会增加评议费用，降低了效率。
3. 分配的专家数量较少而需要评议的材料较多时，就容易出现专家无法在给定时间内完成评议工作，或者专家拒绝参与评议工作。
4. 还容易出现专家熟悉的领域与需要评议的材料不匹配的情况，这将会导致专家拒绝评议，或者出现专家评议质量不高的现场。

如何更加高效、更加公平、更加合理地分配专家则是系统一个痛点，因此有必要研究一种合适的分配专家方法，并实际应用学科平台系统中，主要有以下重要意义：

1. 充分利用现有的计算机系统资源和技术减轻学科平台工作人员的劳动强度，尽量减少学位办相关工作人员分配专家的工作量，让专家自由地、舒心地参与评议工作。
2. 与传统的人工分配专家相比，基于遗传算法的学科平台系统能够有效地确保学科平台的客观性、公正性、实时性和准确度，具有规范管理、提高效率、节约资源等不可替代的优势，不仅减轻了因分配专家所花费的人力和物力，并且突破了时间与空间的

限制大大改善了专家评议工作的效率,节约评议成本,是积极构建节约型政府的必然选择。

3. 考虑到遗传算法具有内在的隐并行性和稳定性,是求解组合优化的常用算法<sup>[8]</sup>。因此,基于遗传算法的专家分配,既针对性强,又能提高分配专家的效率。

### 1.3 国内外研究现状

“同行评议”这个术语,上世纪 60 年代第一次出现在科学出版方面。同行评议发源于英国,后来美国借鉴并发扬了这种评议制度,在学术研究领域内,同行评议制度已经深深地植根于美国的历史、文化和立法中<sup>[9]</sup>。随着科技的飞跃发展和人们对科学知识的深度探索研究,同行评议作为一种客观公正的评议制度,在社会中扮演着非常重要的角色。专家是同行评议活动中最重要的主体,专家分配则是同行评议的一个关键步骤之一,专家的分配是否科学将直接影响到评议结果是否科学、合理。

我国对同行评议的研究和应用相对较晚,且研究范围主要侧重于专家库的完善、评议方式、评议结果处理等方面。专家库建设方面,随着国力的不断增强,专家库中专家的数量越来越多,有着丰富的资源可供国家使用,而且专家自身的指标体系越来越完善。评议方式多样化,有传统的会评、网评、盲评等方式,以及多种评议方式一起使用。项目分类也越来越细化,使得项目的管理越来越精确,更加科学地管理项目。

当前,专家分配的研究主要集中在为某一特定项目分配相同数量的专家,很少研究为一组项目分配相同数量专家的情况,这种按照分组来分配相同数量的专家的情况相较于为单一项目分配相同数量的专家的情况就会复杂得多,当一组项目内的项目差别较大时,那么这就给一般的专家分配方法带来了挑战。各类科技项目聘请同行评议专家大多是根据专家的学科分类,而忽略了专家的学术声誉,使得一流的一线科研人员并没有被优先选做同行评议专家<sup>[10]</sup>,使得他们不能成为真正意义上的专家。

“我们的评议人,我们信得过”,是同行评议的基础和希望。因此,遴选专家无疑是评议的关键所在,是专家客观公正地评议的前提,直接影响着评议结果是否真实地反映了实际情况<sup>[10]</sup>。



## 1.4 本文研究内容及组织结构

本文主要讨论遗传算法在科研项目系统专家分配中的应用。共分为六章，结构安排如下：

第一章为绪论，阐述了本文的研究背景及意义，介绍了专家分配的研究现状。

第二章为基于遗传算法的专家分配方法研究。首先，介绍了遗传算法的基本原理，然后详细地描述了学科平台系统的分配专家要求；接着，结合遗传算法特点和专家分配要求分析了基于遗传算法的专家分配方法，并阐述了基于遗传算法的专家分配方法面对的主要问题；最后介绍了基于遗传算法的专家分配方法的主要问题。

第三章为基于遗传算法的专家分配方法在学科平台系统中的应用。首先主要介绍了学科平台系统中有关专家分配的业务需求，主要包括材料管理、分组管理、专家信息管理、专家遴选、专家评议；接着，根据实际项目需求进行了设计。

第四章为遗传算法在学科平台系统专家分配方法中的应用，首先建立了关于学科平台系统专家分配方法的数学模型；接着根据数学模型详细地描述了遗传算法在专家分配中的应用。

第五章为系统的实现与算法分析。首先，详细地介绍了学科平台系统中主要有关功能的实现，侧重介绍了专家分配功能模块的实现，并结合部分网页截图与核心代码，以便实现过程的描述更加清晰；然后，测试了遗传算法在专家分配方法中的参数；最后，分析了遗传算法在学科平台系统专家分配方法中应用的适用性。

第六章是总结与展望，先阐述了本文完成的主要工作，然后描述了对未来工作的展望。

## 第 2 章 基于遗传算法的专家分配方法研究

### 2.1 遗传算法基本原理

遗传算法是一种模拟物种进化方法的启发式算法，是生成优化和搜索问题的重要解决方案<sup>[12]</sup>。遗传算法（Genetic Algorithm, GA）最初是由美国的 Hooland 教授提出全新的理论，为遗传算法奠定了理论基础，从根本上确保了遗传算法是一个能够用来搜寻最优可行解的过程<sup>[13]</sup>。以达尔文的生物进化论和孟德尔的生物遗传学说为基础，然后仿照生物的进化与遗传过程，并遵循“适者生存，优胜劣汰”的规则，通过选择、交叉、变异等一系列的操作，将求解的问题从初始解逐代地逼近最优解。是一种基于达尔文的自然选择和孟德尔的基因遗传学原理的随机并行搜索算法和寻求全局最优解的高效优化方法<sup>[14]</sup>。遗传算法是一种强大的搜索算法，它利用了生物进化论和自然遗传学的操作通过搜索空间而找到问题的最优解决方案<sup>[15]</sup>。遗传算法具有先天性的稳定性、全局寻优性和隐含并行性<sup>[16]</sup>，因此，被国内外学者广泛研究和应用。近年来，遗传算法在研究和应用都取得了不错的发展<sup>[17]</sup>。最为重要的是，遗传算法为解决复杂系统优化问题提供了一个实用的通用框架，在不同问题的具体领域，都具有很强的稳健性，因此广泛应用于很多领域如：组合优化、智能控制、图像处理等。如表 2.1 所示。

表 2.1 遗传算法应用领域及说明

应用领域	说明
函数优化	主要针对非线性、多模型、多目标的函数
组合优化	巡回旅行商问题、背包问题等
智能控制	线性系统识别，非线性系统识别
生产调度	车间调度、任务分配、生产规划等
图像处理和模式识别	扫描、特征提取、图像分割等
机器学习	人工神经网络的网络结构优化设计
人工生命	进化机器人、人工生命模型、免疫系统

### 2.1.1 遗传算法的基本术语

遗传算法是一种自适应的全局搜索算法<sup>[18]</sup>，借鉴的是生物自然选择和遗传机制，同时也是当代应用数学中的一个重要分支，得到广泛应用。下面将介绍一些遗传算法的术语：

1. 个体：指染色体带有一定特征的实体。待求解问题的每一个可行解都是用个体来表示的。

2. 种群：由个体所组成的集合称为种群<sup>[19]</sup>。种群的大小就是该种群所包含的个体数量，一般情况下，遗传算法的种群大小是不变的。

3. 编码：遗传算法在一般情况下是不能直接处理问题空间的参数，因此必须把它们转换成由具有一定结构的基因而组成的染色体，这种转换操作称为编码<sup>[20]</sup>。常用的编码有符号编码、浮点编码、二进制编码等编码方式。基本遗传算法一般采用固定长度的二进制编码方式表示群体中的个体<sup>[21]</sup>。

4. 解码：从基因型到表现型的映射。即将遗传算法所搜索到的最优解个体转换成问题解形式的过程。

5. 适应度：生物学家用适应度来度量某物种对其所在生存环境的适应能力<sup>[22]</sup>。适应度函数是所求问题的解与遗传算法的桥梁，是种群个体优劣评估的标准<sup>[23]</sup>。适应度值越小的个体就越容易被淘汰，适应度值越大的个体就越容易遗传到下一代。算法的效率与适应度函数设计的优劣息息相关。

6. 选择算子：选择操作又称为复制运算<sup>[24]</sup>，是指按照一定的概率从种群中选择一定数目的个体的操作。通常情况下，选择的过程是基于适应度函数的“适者生存，优胜劣汰”的过程<sup>[25]</sup>。在遗传算法中，有两类的选择：一类是确定选择哪些个体进入到下一代；另一类是选择个体进行遗传操作。基本遗传算法中采用比例选择算子，常见的选择算子方法有最佳保留选择、轮盘赌模型、随机竞争选择等。

7. 交叉算子：将种群内的所有个体随机地两两搭配成对<sup>[26]</sup>，每一配对的两个个体按照一定的概率来交换他们之间的染色体，将两条染色体的某一相同位置处的 DNA 片段切断，其前后两串分别交叉组合形成新的两条染色体<sup>[27]</sup>。在该算法中，交叉方式丰富多样，交叉运算的位置和交叉的基因个数都是随机选择。为了保证算法的收敛速度，要确保交叉生成的新个体仍然是待求解问题的有效解。

8. 变异算子：变异算子是对种群中的部分个体以一定小概率进行变异操作<sup>[28]</sup>。模仿的是染色体进行复制时，染色体以较小的概率产生差错<sup>[29]</sup>，从而导染色体上 DNA 发生变异，这样就产生出新的染色体，新的染色体表现出新的性状。变异的好处是增加了个体的多样性，但是也可能破坏个体中原有的良好的基因。

### 2.1.2 基本遗传算法的操作步骤及流程

遗传算法的主要操作步骤有：确定编码方案、随机初始化种群并设置种群规模大小、计算个体适应度、选择、交叉、变异 6 个基本操作步骤。运算流程如图 2.1 所示。

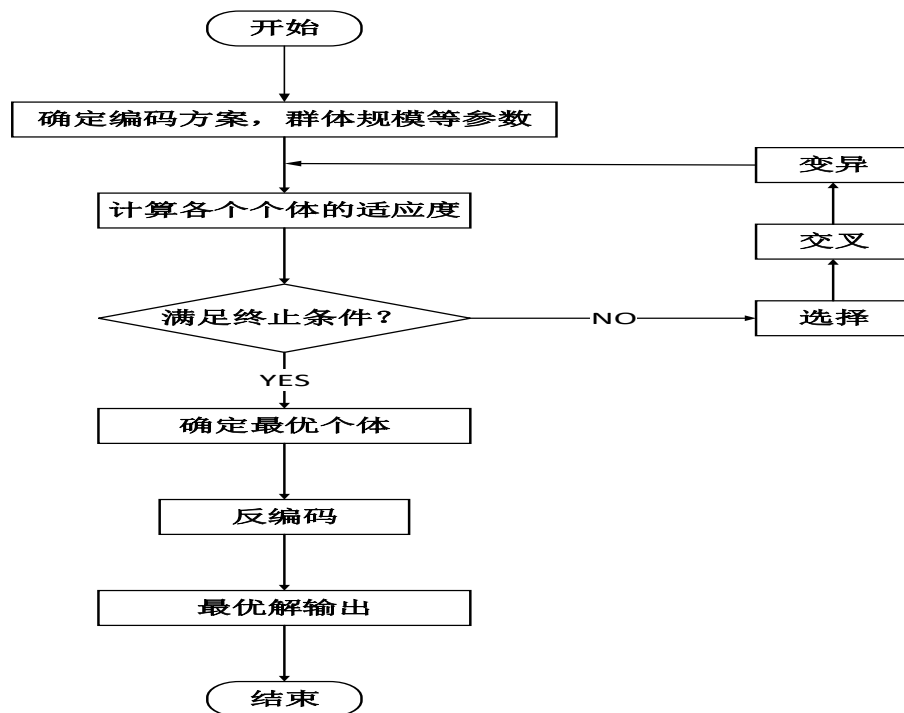


图 2.1 遗传算法运算流程

#### 1. 确定编码方案

编码在遗传算法中占据着一个至关重要的地位，是应用遗传算法的第一步。不同的具体问题对应着不同的编码，编码的好坏将直接影响其他的遗传操作，包括选择、交叉、变异等<sup>[30]</sup>。

#### 2. 随机初始化种群

产生初始种群一般有两种方式。一是完全随机地产生，它适合于对没有任何先验知识的情况。另一种是利用先验知识产生，这种选择方式可以使遗传算法更快地达到最优

解。遗传算法采用随机方式选取  $n$  个个体,并将  $n$  个个体采用设定的编码方案进行编码,这些个体就组成了初始化种群。

### 3. 计算个体适应度

在遗传算法中引入了自然界中生物遗传的适应度概念<sup>[31]</sup>。个体适应度衡量个体距离最优解的程度。因此,需要在遗传算法流程中采用适应度函数来计算种群个体的适应度,适应度值的大小决定了种群每个个体的优劣,也决定了每个个体遗传到下一代的概率。

### 4. 选择

选择概率越大则该个体被选择的概率越大,选择概率越小则该个体被选择的概率越小<sup>[32]</sup>。常见实现方法有轮盘选择、锦标赛选择、最佳选择等,轮盘选择的优点是具有良好的收敛速度。轮盘选择模型中个体被选取的概率与该个体的适应度值成正比关系,适应度值越高,被选中的可能性就越大,进入下一代的概率就越大<sup>[32]</sup>。

### 5. 交叉

交叉运算执行的是全局搜索,是遗传算法中的基本运算<sup>[33]</sup>。选择操作虽然能够从旧种群中选择出优秀者<sup>[34]</sup>,但不能创造新的染色体<sup>[35]</sup>。交叉运算模仿的是生物学中遗传进化过程中的繁殖现象,通过两个染色体的交换组合,来产生新的优良品种。交叉的过程是:在匹配池中任选两个染色体,随机选择一点或多点交换点位置,交换双亲染色体交换点右边的部分,即可得到两个新的染色体数字串<sup>[36]</sup>。交叉运算体现了自然界信息交换的思想。单点交叉的动作过小,搜索的广度不够<sup>[37]</sup>,那么遗传算法的收敛速度大大降低,特别是当问题的复杂度较高时,遗传算法的收敛速度下降得尤其明显,表现在遗传算法所求得解与最优解距离太远。

### 6. 变异

变异运算用来模拟生物在自然的遗传环境中由于各种偶然因素引起的基因突变现象,它以很小的概率随机地改变遗传基因的值<sup>[38]</sup>。若只有选择和交叉,而没有变异<sup>[39]</sup>,则遗传算法就没办法在最初始基因组合以外的空间进行搜索,这就会过早地陷入局部最优解而进入终止过程,从而达不到预期的结果。要想在尽可能大的空间中获得质量较高的解,我们就必须采用变异运算。

变异概率( $P_m$ )能够控制新基因导入群体的比例<sup>[40]</sup>。如果变异概率值过小,那么一些有用的基因就不能进入选择阶段;如果变异概率值过大,意味着随机的变化太多,那

么子代就很可能失去从双亲继承下来的优良特性<sup>[41]</sup>。具体的操作方法可以是：当群体最大适应度值与平均适应度值接近时，则遗传算法趋于收敛阶段，此时应该增大变异概率；反之，则意味着群体的多样性比较强，此时应该减小变异概率<sup>[42]</sup>。

## 2.2 学科平台系统中专家分配方法的研究

分配专家的基本原则是公开、公正、可靠、有效、经济。专家分配是给一组材料遴选指定数量的专家的过程。评议专家的素质及其对所评学科的匹配程度，将直接影响评议结果的质量。在分配专家时，依据系统中设计实现的专家分配方法，系统能够自动地从专家库中遴选出符合要求的专家。分配专家的前提是要确定分配专家的要求和参数，即分配专家的约束条件，例如：专家所属门类、专家所属一级学科或专业学位、所在单位等。分配专家是面向多个目标优化，生成的专家要尽可能满足多个约束条件的要求。

在学科平台系统中常用的分配专家方法有人工方式分配专家的方法、基于回溯算法的专家分配方法、基于遗传算法的专家分配方法，下面对这三种分配专家方法进行对比分析，以便确定本系统采用的分配专家方法的选择。

1. 人工方式分配专家的方法：根据组内材料的特点，人工地分配专家。操作简单，易于实现。但是，分配专家相当繁琐，尤其是专家库中专家数量庞大时，速度很慢，甚至分配专家出现失误。在开发时，主要通过 SQL 语句实现。适用于分组较少，而且专家库中专家较少的情况。

2. 基于回溯算法的专家分配方法：类似于枚举专家的过程，主要是在尝试过程中寻找满足约束条件的专家分配组合，直到找到满足约束条件的任意一个专家分配组合则分配完成。在搜索过程中，对当前节点进行是否满足约束条件的验证，若成功，则记录当前节点，按照深度优先搜索策略，继续往深层搜索，否则，跳过对以该节点为根的子树进行搜索，同时逐层向该节点的祖先节点进行回溯，转向其余未被搜索的节点继续搜索。该方法实际上是对随机分配专家方法的改进，但是，仍旧具有很强的随机性，实际应用中分配专家效率低，分配的质量不高，适用于专家库中专家数量较小的情况。

3. 基于遗传算法的分配专家方法：遗传算法使系统具有出色的自适应性和优化能力<sup>[43]</sup>，是一种自适应全局优化概率搜索算法。上述三种分配专家方法的对比,如表 2.2 所示。

表 2.2 常用分配专家方法对比

方法名称	优点	缺点	应用场景
人工方法	简单、易于实现	分配专家质量、效率低	小容量的专家
回溯算法	通用性较强	多条件约束下分配专家 质量不高	专家数量较少、约束 条件少
遗传算法	求解速度快并且所得 解较优，通用性强	易陷入局部最优	适合所目标、多约束 条件多约束

### 2.3 基于遗传算法的专家分配方法的分析

随着专家评议工作信息化的逐渐完善，学科平台系统也变得越来越成熟，但是，学科平台系统还是存在一些难点和缺陷。比如，专家的分配过程相当繁琐，工作量也较大，专家分配容易出现不合理的现象。针对这些问题，专家分配方法在学科平台系统中的研究变得十分重要，选择合适的专家分配方法，可以有效地实施专家评议工作。基于遗传算法的专家分配方法在解决多目标、多变量、复杂的专家问题中会更加适合。其在学科平台系统中的应用具有以下特点：

1. 黑箱式结构：基于遗传算法的专家分配方法根据专家的特性对专家进行编码(输入)和计算个体适应度(输出)，只考虑输入与输出之间的关系。面向学科平台系统专家分配方法中参数的多样性，该分配方法为此提供了理论基础。

2. 智能搜索寻找全局最优解：基于遗传算法的专家分配方法根据适应度函数智能地高效搜索，并利用选择算子、交叉算子、变异算子等一系列操作<sup>[44]</sup>，产生新的个体，使得下一代结果优于上一代，避免局部最优，最后优化结果逼近全局最优解，该分配专家方法为此提供了可行性。

3. 可行性计算：基于遗传算法的专家分配方法每次迭代过程都是对群体中的所有个体进行运算，覆盖面广，搜索速度快。当学科平台系统中专家规模较大问题时，该分配专家方法为此提供了较高的分配专家效率保障。遗传算法本质上是平行的，因为它的适应度评估和进化过程可以同时进行<sup>[45]</sup>。

4. 通用性强：遗传算法的一个显着优势是它能够直接处理决策变量的离散性<sup>[46]</sup>，传统的优化技术没有这种能力，另外遗传算法不需要设定明确的数学表达式，只需要对

每个约束条件建立简单的原则要求，因此可以应用在函数关系不明确的复杂问题中。面向学科平台系统中多目标优化问题，可以为此抽象出较优的数学模型。

5. 扩展性高：基本遗传算法为遗传算法的应用提供了一个最基本的框架<sup>[47]</sup>，在解决实际问题时，需要根据问题的具体特性而采取具体的方法，比如采用修改编码方案、改善初始化种群、交叉算子等方式使解决问题的效率更高，效果更好。除此之外，遗传算法还易于与其他算法进行结合使用。因此，为基于遗传算法的专家分配方法在学科平台系统应用提供了较大的扩展空间。

根据上述提出的专家分配要求结合分析，本系统拟采用基于遗传算法的专家分配方法进行分配专家。基于遗传算法的专家分配方法在学科平台系统中的分配专家流程如下所示：

1. 确定学科平台系统分配专家中的约束条件，并建立数学模型。设置算法运行时的参数，例如交叉概率、变异概率、算法终止代数、确定种群规模大小、允许的误差值等<sup>[48]</sup>。
2. 按照设定的种群规模大小，初始化种群则由系统随机地初始化生成。
3. 根据适应度函数，对种群中的个体计算其适应度值。若在种群中的最大适应度值的专家满足算法终止允许的误差值，或者种群进化达到终止代数，则算法终止运行。否则就继续循环选择、交叉、变异、更新个体和种群适应度。
4. 选择遗传算子。
5. 若随机概率小于交叉概率，则随机选择专家进行交叉运算。
6. 若随机概率小于变异概率，则随机选择专家，在变异点进行变异运算。
7. 按照以上描述，基于遗传算法的专家分配方法分配专家流程如图 2.2 所示。



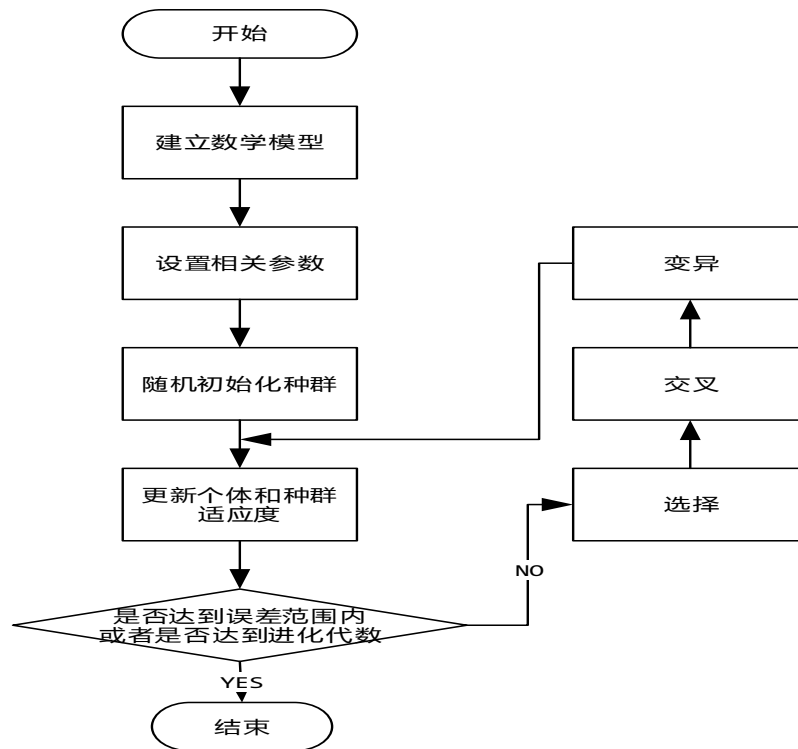


图 2.2 基于遗传算法的专家分配方法流程

## 2.4 本章小结

该章首先简单介绍了遗传算法的基本结构、发展概况以及应用领域，介绍了遗传算法中一些基本的术语；然后，详细描述了遗传算法的完整流程；接着，详细地描述了学科平台系统中专家分配方法的研究，包括：学科平台系统中专家分配的基本要求、学科平台系统中常用分配专家方法的对比分析；接着，依据分配专家要求，结合遗传算法特点，分析了基于遗传算法的专家分配方法，确定了本系统的专家分配方法选择；然后，详细的介绍了基于遗传算法的专家分配方法应用中主要的问题；最后，结合流程图，介绍了基于遗传算法的专家分配流程。

## 第3章 学科平台系统专家分配的需求分析

需求分析，就是要全面地理解用户的各项要求，并准确地表达用户需求<sup>[49]</sup>。通过对学科平台系统的需求进行调研，广泛征求省学位办的意见和建议，经过多次协商讨论，确定系统的总体需求。本章首先对学科平台系统专家分配相关的需求进行描述，并加以分析，然后结合实际的需求对本系统的功能模块的设计进行了详细介绍。

### 3.1 学科平台系统的总体需求

本文旨在建立某省学位与研究生教育管理信息系统平台，实现整个工作流程的一体化管理。运用科技手段，创新工作思路，完善工作措施，确保专家评议过程平稳有序开展。由于系统中其他功能与专家分配相关不大，所以系统的需求是针对专家分配相关的需求。

系统中与专家分配相关的总体需求功能如图 3.1 所示。

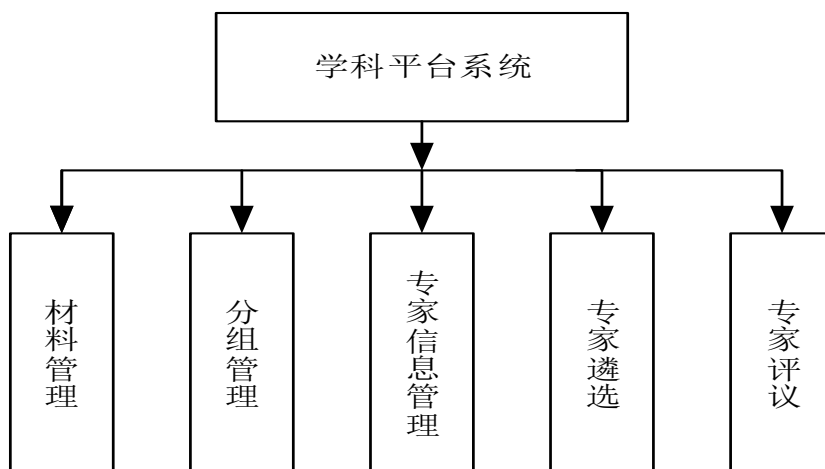


图 3.1 系统总体需求用例

学科平台建设背景是国务院学位委员会开展新一轮博士硕士学位授权工作。各个高校将本校的学位点评审材料电子化，上传到学科平台，省学位办按照材料的门类，一级学科等进行分组，每组聘请固定数量的同行专家进行评审，按照一定规则给评审材料打分，根据专家打分情况，对学位授权审核进行评判。学科平台建设的难点就是如何根据学位点材料情况进行合理分组，并聘请合适的专家对材料进行评审。分配专家的工作不仅在本文的学科平台建设中是很重要的一个环节，在论文盲审，项目评标等系统建设中

也是不可或缺的一个环节。专家的分配的好坏直接影响评审的质量。下面针对材料分组和专家分配的主要需求展开讨论。

1. 组内的申请书材料的一级学科或专业学位很可能不同，比如软件工程与土木工程，虽然两份材料的所属门类都是工学，但是它们的一级学科或专业学位不同。同理，组内的材料的一级学科或专业学位与专家所属的一级学科或专业学位也可以按照这样的比例来遴选。

2. 组内的申请书材料的硕博类型也可能多样，申请书材料的硕博类型有学术型博士、学术型硕士、专业型博士、专业型硕士，这四类硕博类型在一个组内可能都会存在。那么在遴选专家时，尽可能地避免专业型专家去评议学术型材料，硕士生导师去评议博士型材料。

3. 组内的材料的所属单位与专家的所属单位尽可能地避免相同，需要避免专家既是运动员，又是裁判员的现象。

4. 每个组内必须有一名专业教指委主任成员和专业学位教指委委员。

5. 系统的核心是专家管理和材料管理，另外还有基本的用户管理和角色管理，以及系统管理等管理工作，其主要需求包括以下几项：

6. 材料管理。录入学科点申报书信息、验证已录入申报书、申报书资格审查、全部申报书确认提交、统计查询申报书信息。

7. 专家评议。专家评议过程：阅读邀请函和省学位委员会《通知》、评议、网上提交评议结果、打印评议结果并签字、上传或邮寄评议结果。专家可以查询自己的评议进度，可以阅读相关的政策文件和帮助文档，并提供下载。此操作仅限于专家这个角色。

8. 分组管理。材料分组的定义、分组、查询。此操作仅限于材料分组管理员。

9. 专家遴选。专家遴选的过程：给每一组材料遴选指定数目的专家，然后确认并提交专家分组，再确认专家参加评议并创建专家系统管理账号与密码，最后给遴选出来的专家发送邮件。可以查看每组分配的材料和专家。此操作仅限于专家遴选管理员。

10. 评议监控和统计查询。可以按照分组、专家来监控，也可以按照分组、材料来监控。统计查询，可以按照材料来统计查询，也可以按照专家来统计。

## 3.2 学科平台系统详细需求分析

根据系统中与专家分配相关的总体需求可知，由材料管理、分组管理、专家信息管理、专家遴选，专家评议这五部分组成。下面将分别从这五个方面进行详细的需求分析。

### 3.2.1 材料管理

材料是专家分配的基础，是专家分配的主体目标，也是专家分配的依据之一。申请书材料主要有学科门类、一级学科或专业学位、硕博类型、分组名称、申请书材料状态、材料所属单位等属性，专家分配就是根据材料这些特定属性来遴选专家的。

每所申报单位都必须提供 pdf 格式的申请书材料，根据单位提供的申请书材料录入学科点申报书信息，所以接下来需要验证已经录入的学科点申报信息，然后需要对每个学科点申报信息进行资格审查，材料经过资格审查后才能对材料进行分组，可以对申请书材料进行统计和查询。

对申请书材料进行分组主要是依据申请书材料的学科门类、一级学科或专业学位、硕博类型，相似学科尽量分在一组，另外还可以对每个组中的材料都可以进行新增、删除、修改、查看操作。对每个组可以进行新增、删除、修改、查看操作，这些操作需要特别慎重，因为后续的操作都需要在这个分组的前提下进行，需要有专门的一个角色进行材料分组操作。

根据上述需求分析可知，材料管理需要的功能如图 3.2 所示。

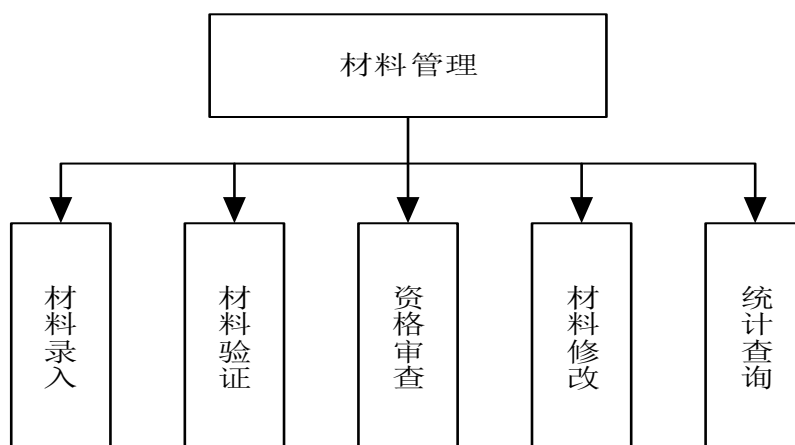


图 3.2 材料管理功能

### 3.2.2 分组管理

分组是材料和专家之间的桥梁，分组的大小是根据材料的多少而定的，分组不仅仅包括定义分组、修改分组、删除分组、查询分组这些基本功能，而且还要给所有定义的分组筛选材料。专家的分配是基于材料分组的，分组中的材料所属一级学科尽可能地相同或者类似，对于分组的修改与删除操作需要特别小心，因为这些操作牵一发而动全身，而且这种操作特别敏感，操作仅限于材料分组管理员。

根据上述需求分析可知，材料管理需要的功能如图 3.3 所示。

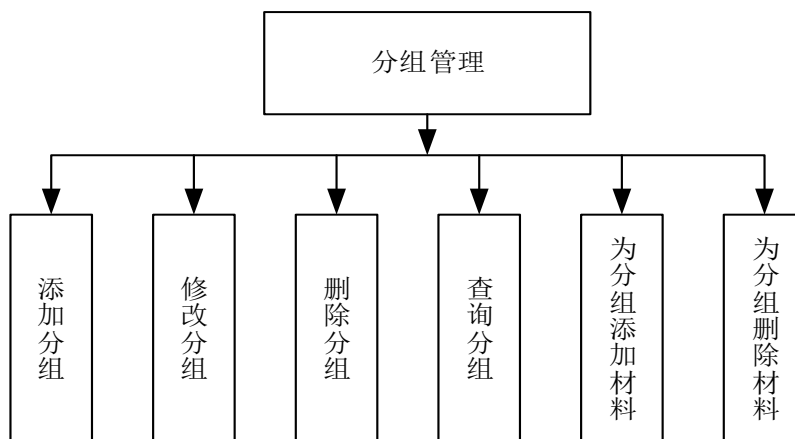


图 3.3 分组管理功能

### 3.2.3 专家信息管理

专家是整个系统的主体对象，专家信息管理主要是维护专家信息的工作，专家库的资料由教育厅提供，需要将信息录入系统，专家的信息有可能随着时间而变动，即需要提供专家信息的维护工作，确保专家信息准确无误是基本的要求。

专家管理模块的基本功能包括专家库的维护。其中，专家库的维护包括：专家的查询、录入、修改、删除等。

根据上述需求分析可知，专家管理需要的功能如图 3.4 所示。

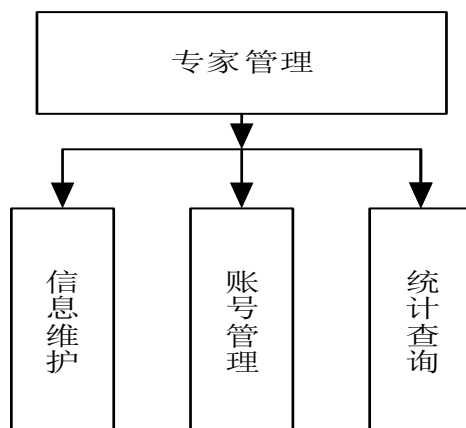


图 3.4 专家管理功能

### 3.2.4 专家遴选

专家的遴选工作是在材料分组完成后进行的，给每一组材料分配指定数量的专家不是随机选择，需要按照一定的规则遴选的，如果采用人工遴选的话，就会存在人为因素过多，人工遴选存在不知道遴选谁的尴尬，为了公平起见，故采用良好的规则进行机选是必要的。材料主要有学科门类、一级学科或专业学位、硕博类型、所属单位、推荐申报级别这些关键属性，专家则有学位、是否学科评议组成员、是否学科评议组召集人、是否专业学位教指委委员、是否专业教指委主任委员、是否省级学位委员会委员、是否博士学位授权高校的校领导、是否博士学位授权高校校长、学术学位研究生导师类别、专业学位研究生导师类别、现任专技职务、行政职务等属性，如何给一组材料分配指定数量的专家需要解决。

分配专家的主要功能是，专家评议工作人员系统管理系统后，可以根据实际的专家评议场景，选择手工分配和基于遗传算法的智能分配两种方式进行分配专家，手工分配专家时，专家评议工作人员指定分配专家条件，系统随机从材料中生成专家，专家分配时，根据专家评议工作人员设置的约束条件，系统自动的从材料中搜索，生成满足约束条件的专家，分配专家完成后，专家评议工作人员需要为专家添加必要的参数信息，保证专家的完整。专家评议对象在进行学科平台时，专家出现在专家中的位置均是随机的，这样可以防止作弊现象，保证专家评议效果的真实性。

### 3.2.5 专家评议

参与评议的专家必须从专家库中选择,专家需要阅读邀请函和省学位委员会《通知》,然后根据每个学科点的具体要求评议每一份申请书材料,当专家评议完所有分配给自己的材料后需要网上提交,然后需要打印评议结果并签字,最后需要上传或者邮寄评议结果,专家可以随时查看自己的评议进度,可以阅读相关的政策文件,系统需要提供一些帮助文档,以方便专家评议,界面需要干净简洁。

专家评议过程需要提供监控。评议过程是需要一段时间才能完成的,考虑到有的专家工作繁忙,不一定会及时地评议,这时就有必要地提醒一下。专家评议过程可以分两个角度监控,一是根据分组、材料,这样先查看某一分组的整体评议进度,还可以深入到一个分组内每一份申请材料的评议进度,另一个角度是根据分组、专家,可以查看某一个分组的整体评议进度,还可以深入到一个分组内每位专家的评议进度。

专家评议模块负责专家评议工作的相关功能。评议过程是:

1. 需要阅读邀请函和省学位委员会《通知》。
2. 根据每个学科点的具体要求评议每一份申请书材料。
3. 网上提交。
4. 打印评议结果并签字。
5. 上传,或者邮寄评议结果。

专家评议模块的主要步骤如图 3.5 所示。

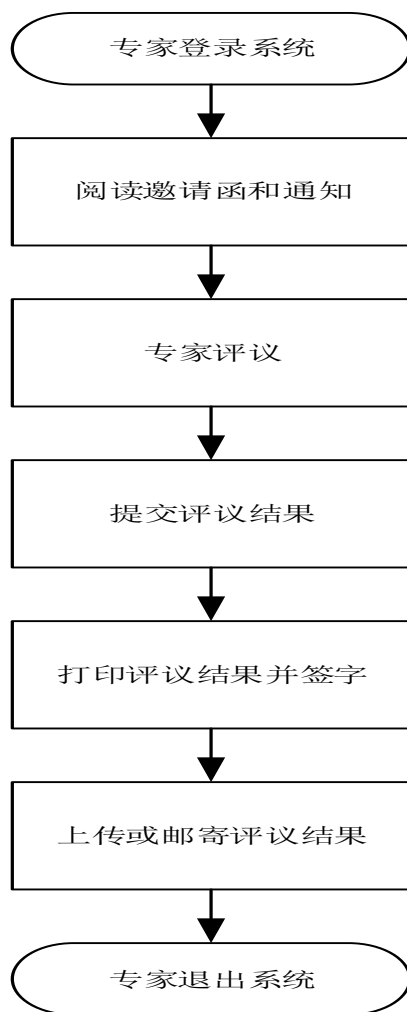


图 3.5 专家评议步骤

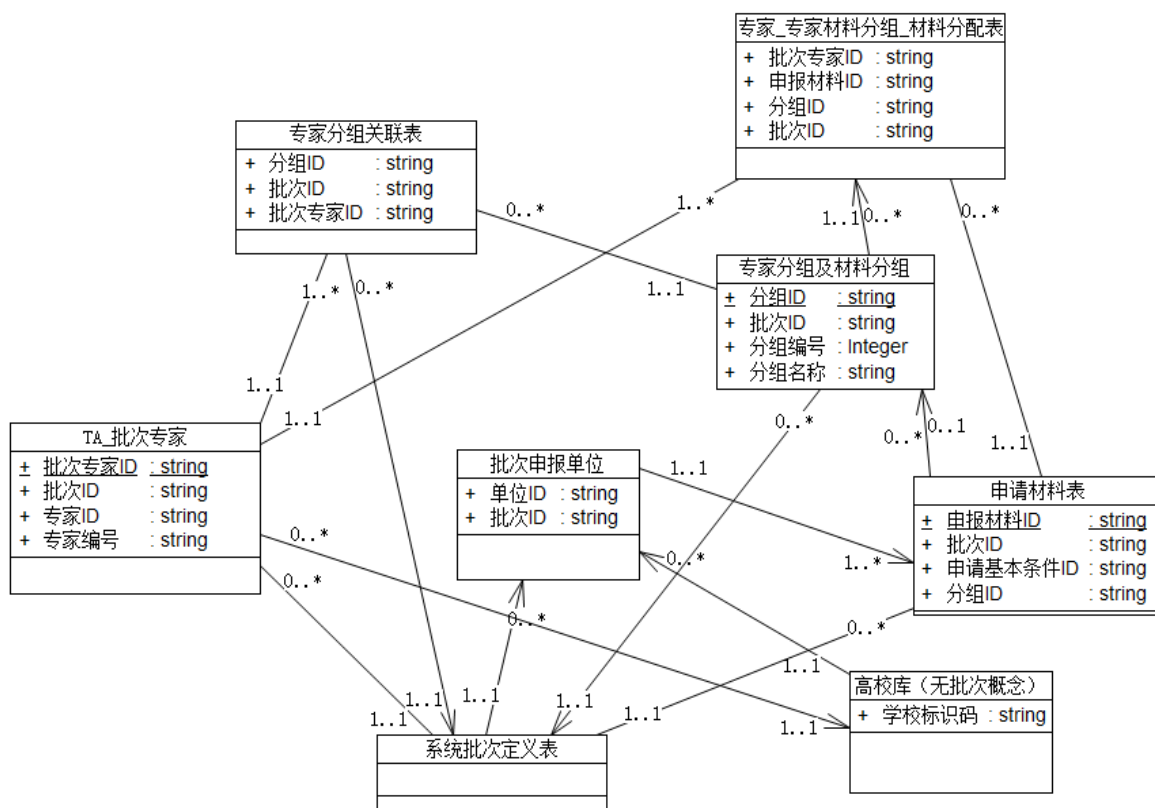
### 3.3 学科平台系统设计

根据前面需求分析之后，先简单介绍与专家分配相关的类图设计，然后简要介绍与专家分配相关的表设计。

#### 3.3.1 与专家分配相关的类图设计

本系统中类比较多，为了便于描述，进行了简化处理，将系统中的代码表、高校库等均已省略，本小节只对与专家分配相关的核心类进行了展示，核心类主要是申请书材料类、专家类、专家分组及材料分组类等，如图 3.6 所示。





### 3.3.2 与专家分配相关的表设计

本小节将简单地介绍申请书材料表、专家表、专家分组及材料分组表，这三张表与专家分配息息相关，而且每张表只列出了与专家分配密切相关的列。

表 3.1 申请书材料表

列名	数据类型	说明
材料编号	VARCHAR2(32)	主键
一级学科名称	VARCHAR2(32)	
一级学科代码	VARCHAR2(11)	
学科门类名称	VARCHAR2(256)	
学科门类代码	VARCHAR2(8)	
所属单位名称	VARCHAR2(256)	

专家表，主要列出了专家编号、一级学科、学科门类、信誉等级等列属性，这些属性是遴选专家的重要依据。如表 3.2 所示。

表 3.2 专家表

列名	数据类型	说明
专家编号	NUMBER	主键
一级学科名称	VARCHAR2(20)	
一级学科代码	VARCHAR2(11)	
学科门类名称	VARCHAR2(256)	
学科门类代码	VARCHAR2(11)	
所属单位名称	VARCHAR2(256)	
所属单位代码	VARCHAR2(11)	
信誉等级	CHAR(1)	

专家分组及材料分组表，主要列出了分组编号、分组名称、批次编号、等列属性，这些属性也是比较重要的。如表 3.3 所示。

表 3.3 专家分组及材料分组表

列名	数据类型	说明
分组编号	NUMBER	主键
分组名称	VARCHAR2(32)	
批次编号	VARCHAR2(16)	
专家分组状态代码	VARCHAR2(8)	用于专家状态标识
专家分组状态名称	VARCHAR2(32)	同上

### 3.3 本章小结

本章首先介绍了学科平台系统中与专家分配相关的需求，然后按照系统材料管理、分组管理、专家信息管理、专家遴选、专家评议这五个主要模块进行了详细的需求介绍，最后，在描述分析过程中，引入了功能图、类图等让系统的需求分析与设计过程更加清晰。

## 第4章 基于遗传算法的专家分配方法在学科平台系统中的应用

在第二章中研究了基于遗传算法的专家分配方法，并结合第三章介绍了学科平台系统专家分配的需求分析与设计，本章首先根据实际需求建立专家分配的数学模型，然后根据学科平台系统专家分配的要求逐步确定遗传算法中的核心要素。将遗传算法应用于学科平台系统的专家分配方法中，满足专家评议需求，同时考虑尽量提高分配专家效率。应用遗传算法的流程如图4.1所示。

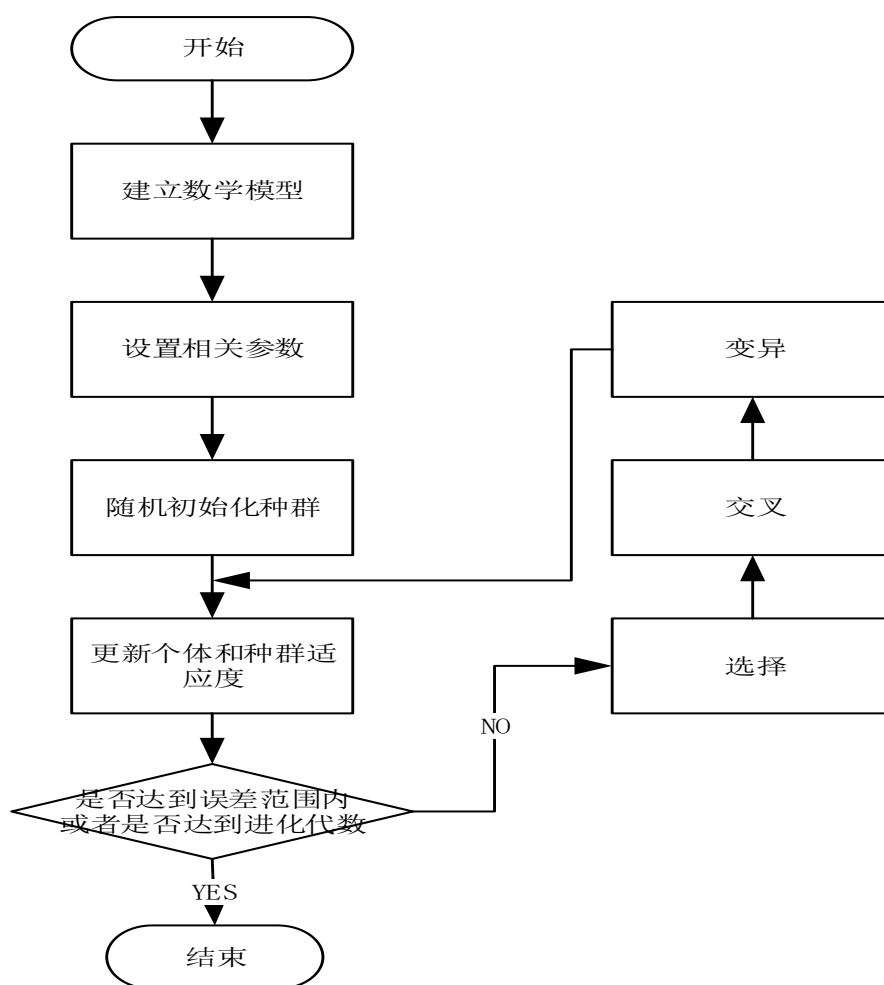


图4.1 专家分配中的遗传算法流程图

### 4.1 专家分配问题的数学模型

专家分配数学模型的建立是算法实现的基础，模型建立的优劣程度将直接影响算法

的运行效率,分配专家的效果。因此,在详细描述了学科平台系统的需求后,将结合实际需求,建立该系统的专家分配数学模型。

#### 4.1.1 专家分配的问题描述

为了便于描述专家分配问题,作了一些必要的假设。假设将所有材料分组成  $n$  个分组,每一分组内的材料的所属一级学科或专业学位可能都相同,也可能不同,然后从专家库中遴选出  $m$  位专家,要求是给每一分组遴选出相同数量的专家,这里假设每组分配的专家数量为  $a$ ,此问题可以用下列矩阵表示:

$$x_{ij} = \begin{cases} 0 & \text{当专家 } j \text{ 与分组 } i \text{ 没有评议关系时} \\ 1 & \text{当专家 } j \text{ 与分组 } i \text{ 有评议关系时} \end{cases}$$

矩阵中行编号  $i$  表示分组编号,列编号  $j$  表示专家编号,因此  $x_{ij}$  表示专家编号为  $j$  的专家与分组编号为  $i$  的分组是否有评议关系,若有评议关系,则  $x_{ij} = 1$ ,否则  $x_{ij} = 0$

下面举例说明,例如将 7 位专家分配给 3 个分组,每组分配 3 位专家,分组编号分别是  $g_1$ 、 $g_2$ 、 $g_3$ ,专家编号,分配如表 4.1 所示。

表 4.1 专家分配示例表

专家 分组	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$
$g_1$	0	1	1	0	0	1	0
$g_2$	0	0	1	1	0	0	1
$g_3$	1	0	0	0	1	0	1

矩阵中行编号表示分组编号,列编号表示专家编号,上表的含义是将  $e_1$  专家分配给了  $g_3$  分组,  $e_2$  专家分配给了  $g_1$  分组,  $e_3$  专家分配给了  $g_1$  和  $g_2$  分组,  $e_4$  专家分配给了  $g_2$  分组,  $e_5$  专家分配给了  $g_3$  分组,  $e_6$  专家分配给了  $g_1$  分组。  $e_7$  专家分配给了  $g_2$  和  $g_3$  分组。

### 4.1.2 材料和专家的处理

为了精确并且高效的完成分配专家过程，需要在满足经典测量理论的基础上结合具体的实际需求来设置专家的相关属性。专家的属性过多，会导致分配专家效率变低，往往也不容易找到最优解，专家的属性过少，就不能满足实际需求，分配专家的目的达不到。

材料和专家是分配专家的核心要素，材料和专家的属性较多，为了便于处理，适当地进行简化，则有利于问题的解决。材料和专家的属性处理。材料的学科门类、一级学科或专业学位、所在分组和材料所属单位是核心属性，专家的编号、一级学科或专业学位、学科门类和信誉等级是专家的核心属性。

专家的信誉等级处理。信誉等级按照等级从高到低划分为 A、B、C、D、E。为了便于计算，将信誉等级对应量化为数字是可行的处理，因此分别将 A、B、C、D、E 对应到数字 5、4、3、2、1。这就将信誉等级转化成遗传算法能够处理的数据。信誉等级量化处理如表 4.2 所示。

表 4.2 信誉等级量化处理表

信誉等级	量化处理
A	5
B	4
C	3
D	2
E	1

专家与分组之间的学科紧密程度处理。一个材料分组可能存在着不同一级学科或专业学位，甚至是不同的学科门类，因此要求一位专家具有材料分组中存在的所有一级学科或专业学位是不现实的，因此将专家的一级学科与材料分组中最相关的一份材料作为该专家与该分组材料的学科紧密度是合理的，那么问题就简化为专家与分组中关系最紧密的那一份材料作为专家与一个分组的整体学科紧密度。分组中的材料可能存在专家的一级学科或专业学位，也可能只存在学科门类相同，还可能学科门类都不同。为了便于量化，如果该专家的一级学科或专业学位在该分组中存在对应的一级学科或专业学位的

材料, 则该专家与该分组的学科紧密度为 4; 如果该专家的一级学科或专业学位在该分组中不存在相同的一级学科或专业学位, 而只有相同学科门类的材料存在, 则该专家与该分组的学科紧密度为 2; 如果该专家的学科门类与该分组中所有材料所属的学科门类都没有相同的, 则该专家与该分组学科紧密度为 1。专家与分组之间的学科紧密度量化处理如表 4.3 所示。

表 4.3 专家与分组之间的学科紧密度量化处理表

专家与分组的紧密度	量化处理
存在一级学科相同	4
仅存在学科门类相同	2
学科门类不相同	1

专家与分组之间的总紧密处理。这里简化处理为专家的学科紧密度与信誉等级之间的乘积作为专家与该分组之间的总体紧密度。举例说明, 现假设一共 7 位专家, 依次是  $e_1, e_2, \dots, e_7$ , 一共有 3 个分组, 依次是  $g_1, g_2, g_3$ , 现将这 7 位专家分配给这 3 个分组, 每组要求分配 3 位专家, 分配结果如表 4.1, 若  $e_1$  专家与分组  $g_3$  之间的学科紧密度为 4, 信誉等级为 3;  $e_2$  专家与分组  $g_1$  之间的学科紧密度为 2, 信誉等级为 5,  $e_3$  专家与分组  $g_1, g_2$  之间的学科紧密度分别为 1 和 4, 信誉等级为 3;  $e_4$  专家与分组  $g_2$  之间的学科紧密度为 1, 信誉等级为 5,  $e_5$  专家与分组  $g_3$  之间的学科紧密度为 2, 信誉等级为 4,  $e_6$  专家与分组  $g_1$  之间的学科紧密度为 4, 信誉等级为 5,  $e_7$  专家与分组  $g_2, g_3$  之间的学科紧密度分别为 1 和 4, 信誉等级为 5, 上面没有列出专家对每个分组的学科紧密度量化值, 是因为由表 4.1 的专家分配结果得知, 计算专家与无关分组之间的学科紧密度没有意义。则由上面的数据可得知, 专家与分组的总紧密度为下表 4.4 所示。

表 4.4 专家与分组之间的总紧密度表

专家 分组	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$
$g_1$	0	10	3	0	0	20	0
$g_2$	0	0	12	5	0	0	5
$g_3$	12	0	0	0	8	0	20

### 4.1.3 分配专家的约束条件确定

分配专家过程中，除了专家与材料的门类、一级学科等属性需要基本匹配外，还有一些其它的约束条件，下面对这些约束条件是在专家分配过程中必须考虑的。

1. 每个分组都有相同数量的专家。即便一个分组只有一份材料，那么也需要分配相同数量的专家，另外分配专家的数量必须是奇数。

2. 专家所属单位与分组中的所有材料的所属单位都不能相同，即回避专家评议自己所在单位的情况。

3. 每位专家评议能够评议的分组数目最多不超过  $c$ ，评议的分组数目不低于  $b$ ，这拓展了系统的适用范围，如果不加这个约束条件，那么就有可能出现一位专家分配到了所有分组中的情况，所以设置了专家能够评议的最大分组数目。另外，考虑到系统有可能会被用到其他评议项目，为了有效利用专家这样的稀缺资源，就有可能要求每位专家至少评议多个分组等情况，因此添加了专家评议的分组数目不低于  $b$  这样的约束条件。

4. 每个分组内的材料的所属一级学科或专业学位的比例与每个分组所分配的专家的所属一级学科或专业学位的比例尽可能接近。这里假设一种情况，假设某一个分组内一共有 10 份材料，其中一级学科为软件工程的材料数量为 8 份，一级学科为翻译的材料数量有 2 份，需要给该组分配 5 位专家。现在，如果给该组分配了 4 位一级学科为翻译的专家，分配了一位一级学科为软件工程的专家，那么这是及其不科学的分配方案，因此要尽可能地将 5 位专家按照材料所属一级学科的比例来分专家，即软件工程与翻译的比例为 4:1，那么相应地专家数量要尽可能地接近 4:1。因此一级学科或专业学位和学科门类都应尽可能地满足比例分配。

#### 4.1.4 建立专家分配问题的目标函数

专家分配是专家评议的基础工作。具体实施时，要综合、均衡地考虑分组与专家的各种因素和关联，为每个分组的材料分配最佳匹配的专家，确保整个分配方案最佳。

根据 4.1.1、4.1.2 和 4.1.3 节，我们可以知道，分配专家的过程其实就是在尽可能满足约束条件的情况下，生成合理的专家，本质上来说，是一种面向多重目标优化的问题<sup>[50]</sup>。具体建立专家分配问题的模型如下：

为了便于描述问题，假设有  $n$  个分组  $g_i$  ( $i=1, 2, \dots, n$ )，有  $m$  位专家  $e_j$  ( $j=1, 2, \dots, m$ )，每个组分配  $a$  个专家。每位专家评议的分组数在  $[b, c]$  范围内。分组与专家之间的评议关系的紧密程度用  $f_{ij}$  来衡量。

$$\max z = \sum_{i=1}^n \sum_{j=1}^m f_{ij} x_{ij} \quad (4.1)$$

$$\min d = \sum_{j=1}^m \text{sign} \left( \sum_{i=1}^n x_{ij} \right) \quad (4.2)$$

$$\min p = \sum_{j=1}^m \text{punish} \left( \sum_{i=1}^n x_{ij} \right) \quad (4.3)$$

$$\min q = \sum_{i=1}^n \text{reduce}(|x_i|) \quad (4.4)$$

$$\sum_{j=1}^m x_{ij} = a \quad (4.5)$$

说明如下：

1. 其中  $\text{sign}(x)$  是符号函数，即当  $x > 0$  时， $\text{sign}(x)$  等于 1；当  $x < 0$  时， $\text{sign}(x)$  等于 -1；当  $x = 0$  时， $\text{sign}(x) = 0$ 。

2.  $\text{punish}(x)$  函数，当  $0 < x \leq b$  时， $\text{punish}(x) = b - x$ ，表示当专家评议的分组数少于最低分组数的要求；当  $x \geq c$  时， $\text{punish}(x) = x - c$ ，表示当专家评议的分组数大于最高分组数的要求；当其他情况时， $\text{punish}(x) = 0$ ，表示当专家评议的分组数合理，因此不用惩罚。

3.  $x_{ij}$  含义，当  $g_i$  与  $e_j$  有评议关系时， $x_{ij} = 1$ ；当  $g_i$  与  $e_j$  没有有评议关系时， $x_{ij} = 0$ 。

4.  $z$  表示分组与专家之间的紧密程度，该值越大表明该专家与所在分组越匹配。



5.  $d$  表示使用的专家数量，该值越小则说明越节省开支。
6.  $p$  表示每位专家参与评议的分组数与最优范围的差距，该值越小越好。
7.  $q$  表示每个分组中材料的一级学科或专业学位的比例与分配的专家所属一级学科或专业学位的比例的差值的绝对值的求和。该值越小表示专家分配的比例越好。

## 4.2 遗传算法在专家分配方法中的应用

遗传算法具有很强的扩展性，虽然基本遗传算法为遗传算法的应用提供了基本的框架。但是，在实际解决问题时，需要对遗传算法进行适当的修改。

根据上述情况，在将遗传算法应用在学科平台系统专家分配方法中时，在遗传算法的基本应用原理上，结合系统特点，通过选取合适的编码方案、改善随机化初始化种群、精英保留方法、优化交叉运算、多种终止条件结合使用等尽量使多约束条件下专家达到最优解的效果，满足专家评议需求。

### 4.2.1 确定编码方案

遗传算法的第一步是编码问题，染色体的编码对算法的性能有很大影响<sup>[51]</sup>。遗传算法一般不直接对实际变量进行操作，遗传算法通过编码来描述问题的可行解，编码方案是遗传算法设计的一个关键步骤，这决定了遗传算法操作算子的作用方式，另外这也决定了遗传算法的求解精度和搜索复杂度。因此，根据具体问题的不同，编码方案的设计或选择也不相同，恰当的编码方案有助于提升算法性能和效率。编码方案多种多样，在遗传算法的应用中用到最多的编码方案就是实数编码和二进制编码。

专家分配问题，为了反映专家分配问题和方便遗传操作，故采用实数编码。具体编码方案如下：

假设有  $n$  个分组，每组分配  $a$  位专家，那么编码的长度  $L$  便可以确定为  $n \times a$ ，编码格式可为：

$g_1$  分组分配的专家为:  $e_{11}e_{12} \dots e_{1a}$

$g_2$  分组分配的专家为:  $e_{21}e_{22} \dots e_{2a}$

.....

$g_n$  分组分配的专家为:  $e_{n1}e_{n2} \dots e_{na}$

那么一个个体染色体可表示为 $e_1, e_2, \dots, e_i, \dots, e_{na}$ , 即一个染色体上共有 $n \times a$ 个基因位。令 $e_i = j$  ( $j = 1, 2, 3, \dots, m$ ), 代表的是专家的编号(整数表示)。这种编码方式就保证了每个分组都有相同数目专家的约束条件, 但仍然要避免一位专家在一个分组内出现多次的情况。

举例说明, 比如有两个分组依次 A、B, 需要为每组分配 3 位专家, 专家库一共有 8 位专家, 专家编号为 1、2、3、4、5、6、7、8, 那么一个染色体可能为 1、2、3、2、5、8, 这表示给“A 组”分配了编号为 1、2、3 的专家, “B 组”分配了编号为 2、5、8 的专家。但是不能出现 1、2、2、2、5、8 这种染色体, 因为这表示给“A 组”分配两位相同专家, 这显然是不合理的, 因此需要避免这种情况的出现。

### 4.2.2 群体规模与初始化

在完成编码方案的设计后, 首先要确定的是群体的大小, 在遗传算法中群体规模的大小对算法的效率和质量有着直接的影响。规模过小, 容易陷于局部最优解; 规模过大, 算法运行时间较长。在遗传算法的应用中, 一般建议将群体规模设置为 20 到 160 之间, 具体的种群规模大小需要结合实际情况而定, 需要实际的测试方可确定。

种群的初始化是算法的开始, 初始种群一般都是随机地产生, 因此初始种群一般也达不到我们的预期, 尤其是数据量庞大的时候。如果初始种群越好, 适应度越大, 那么遗传算法的效率将会大大提高; 如果初始种群较差, 那么遗传算法就会付出更长的时间。

下面将举例说明, 假设种群规模为 3, 分组数量为 2, 每组分配的专家数目为 3, 种群规模为 4, 专家编号为 1 到 100, 设种群中的个体分别为个体 A、个体 B、个体 C、个体 D, 则可能的初始化种群为:

个体 A: 40 55 60    23 45 65

个体 B: 40 14 20    13 35 44

个体 C: 11 19 91    24 37 18

个体 D: 70 13 17    66 88 99

### 4.2.3 适应度函数的设计

适应度是遗传算法中衡量个体的优劣程度的主要指标, 一般地, 个体适应度值越大,

表示该个体适应能力越强，反之，个体适应度值越小，该个体适应能力越弱。遗传算法根据个体适应度值的大小来表达“适者生存，优胜劣汰”的自然法则。通常情况下，适应度函数是由目标函数转化而来的，专家分配是一个多目标优化问题。

对于多目标优化问题，通常将多目标问题转化为称为标量化的单目标优化问题<sup>[52]</sup>，因此在这里采用该方法来处理多目标函数，该方法可描述为：对多目标优化问题的  $m$  个目标按照其重要程度给以适当的权重系数  $w_i > 0$  ( $i=1, 2, \dots, m$ ) <sup>[53]</sup>，且  $\sum_{i=1}^m w_i = 1$  然后求解线性加权和  $\min \sum_{i=1}^m (w_i f_i(x))$ ，用该方法求得的解是原目标优化问题的一个 pareto 解。但是运用该方法时，需要对目标函数做一些一致化和规范化处理，具体处理如下：

$$z^u = \frac{(\max z - z)}{\max z} \quad (4.6)$$

$$d^u = \frac{d}{\max d} \quad (4.7)$$

$$p^u = \frac{p}{\max p} \quad (4.8)$$

$$q^u = \frac{(\max q - q)}{\max q} \quad (4.9)$$

使用加权系数法，可得：

$$\min l = w^1 z^u + w^2 d^u + w^3 p^u + w^4 q^u \quad (4.10)$$

其中  $w^1 + w^2 + w^3 + w^4 = 1$

#### 4.2.4 遗传选择、交叉和变异

##### 1. 选择算子

选择算子。遗传算法的在对个体的适应度评价基础上进行选择操作，个体遗传到下一代的概率与该个体的适应度是成正相关的关系。选择算子的目的是以免避免优良个体丢失、防止陷入局部最优。

其中，最常用的就是轮盘赌模型和精英保留方法<sup>[54]</sup>，在轮盘赌模型中个体被选择的概率与其适应度值大小成正比，但是，由于轮盘赌模型具有随机性，在选择上容易造成误差，所以，有可能导致丢失优良个体，造成平均适应度值下降。精英保留方法，当代种群中适应度最高的那个个体直接跳过交叉和变异运算，直接替换经过交叉和变异运算后产生的适应度最低的个体，好处是保证了优良个体不被丢失，坏处是某个局部最优个体不仅不被淘汰，反而一直遗传。为了避免单纯使用轮盘赌模型或精英保留方法带来的

困扰，因此，分配专家方法中采用基于轮盘赌模型结合精英保留方法来进行选择操作。

在实现过程中，首先，根据个体适应度值的大小，找出并暂存当前种群中个体适应度值最大的个体  $P$ ，然后，对当前种群的所有个体采用轮盘赌模型进行随机选择个体遗传到子代，最后，比较个体  $P$  与子代种群中的个体适应度值作比较，若大于下一代种群中的最大适应度值，则替换掉下一代种群中与最大适应度值个体题型相同且适应度最低的个体，否则，不需要进行替换。

举例说明，假设个体的编码与对应的适应度值如表 4.5 所示。

表 4.5 个体与个体适应度表

个体编码	个体适应度
12 31 77 88 21 66	2
34 22 76 43 89 65	4
33 17 19 44 90 58	3
99 68 24 39 71 81	1

则个体的选择概率为该个体与种群中所有个体适应度之和的比值，为了便于表示个体编码。将上述表格中的个体编码从上到下依次记为 1、2、3、4，则有：

$$P_1 = \frac{2}{10} = 0.2$$

$$P_2 = \frac{4}{10} = 0.4$$

$$P_3 = \frac{3}{10} = 0.3$$

$$P_4 = \frac{1}{10} = 0.1$$

累积概率为个体选择概率与该个体前面的所有个体的选择概率之和，则有：

$$Q_1 = P_1 = 0.2$$

$$Q_2 = Q_1 + P_2 = 0.6$$

$$Q_3 = Q_2 + P_3 = 0.9$$

$$Q_4 = 1$$

个体选择概率和累积概率如表 4.6 所示：

表 4.6 个体选择概率与累积概率表

个体编码	个体适应度	选择概率	累积概率
12 31 77 88 21 66	2	0.2	0.2
34 22 76 43 89 65	4	0.4	0.6
33 17 19 44 90 58	3	0.3	0.9
99 68 24 39 71 81	1	0.1	1

采用轮盘赌模型方法来选择个体，则在 $[0,1]$ 区间内随机均匀地产生四个随机数，这四个随机数假设为 0.55、0.73、0.12、0.86，则表示选中的个体编码如表 4.7 所示。

表 4.7 个体编码表

随机数	选中的个体编码
0.55	34 22 76 43 89 65
0.73	33 17 19 44 90 58
0.12	12 31 77 88 21 66
0.86	33 17 19 44 90 58

## 2. 交叉算子

交叉遗传算子首先以一定的概率发生<sup>[55]</sup>，并产生两个后代。交叉运算是为了保证遗传算法产生更优良的个体，随机选取两个染色体配对，交换部分染色体。由于前面确定了遗传算法的编码方法，个体染色体可表示为  $e_1, e_2, \dots, e_i, \dots, e_{na}$ ，如果允许每个基因位自由地进行交换的话，那么就很可能出现一个分组内分配了两个相同的专家，而这种情况是需要避免的，因此，我们需要采用离散重组的方法，即需要将连续的  $a$  位专家当作整体，并且  $a$  位连续专家的开始位置分布是以 1 位首项， $a$  为公差的等比数列。这就保证了整个染色体编码的有效性。

离散重组是一种交叉算子方法，子代个体的每个基因按照等概率地随机地从父代个体选择，下面则是一种可能的子代，如表 4.8 所示。

表 4.8 均匀交叉算子示例表

个体名称	个体编号
父代个体 P1	2 11 8 9
父代个体 P1	3 5 234 4
子代个体 S1	2 11 8 4
子代个体 S2	2 5 234 9

### 3. 变异算子

变异操作丰富了种群的多样性，当种群趋于局部最优解时，通过适当的变异，可以使遗传算法接近最优解<sup>[56]</sup>。前面编码方式采用的是实值方式，分配专家方法中的变异操作需要将一组内的  $a$  位专家当作整体来操作，即对每  $a$  位专家进行变异操作，新的等位基因随机生成，但是需要排除该染色体中的  $a$  串上的等位基因。

举例说明，假设专家编号从 1 到 100，有 3 个分组，每组分配 3 位专家，那么在进行变异操作时，首先是每个个体按照一定的变异概率发生，倘若一个个体为“20，45，67；10，20，90；89，34，67”的个体恰好命中而发生变异，那么这个个体在具体的那一点发生变异也是按照概率而发生的，一般某个具体点发生变异时，是按照等概率而选择其中一点发生的。假设这个个体在第二段“10 20 90”的第三个等位基因处发生变异，即“90”这个点发生变异，那么“90”发生变异，则需要从 1 到 100 中剔除 10 20 90 这三个编号后的专家中任意选择一个编号来替换掉该“90”，倘若新的基因值为 33，则新的个体为“20，45，67；10，20，33；89，34，67”。如下表 4.9 所示：

表 4.9 单点变异算子示例表

个体名称	个体编码
父代个体 P	20 45 67 10 20 90 89 34 67
子代个体 S	20 45 67 10 20 <b>33</b> 89 34 67

在实现过程，每个基因都有发生变异的概率，在应用过程中，变异概率一般 0.001 到 0.1 之间。

### 4.2.5 遗传算法终止

遗传算法在进化中的过程中一步一步的逼近最优解，但是，在实际情况中想要完全

达到最优解也几乎不可能，如果进化代数过多，算法效率过低，甚至有可能将最优解丢失，所以对遗传算法的终止条件一般有两种方式，一是，设定遗传算法进化终止代数，即当遗传算法进化达到设定值时，算法运行结束，一般取值在 100 到 2000，二是，设定终止进化的阈值，即对个体适应度的期望，当个体适应度达到或逼近这个期望值时，算法结束，此时，认定取得最优个体就是问题的最优解。

在本系统的专家分配方法中，采用将两者结合的方式。如果，算法达到了设定的终止代数，程序运行结束，此时的最优个体就是最优解，即分配专家方案；如果，算法还未运行到终止进化代数，但是，种群中最大适应度值已经逼近设定的期望值，即实际最大适应度值与目标适应度值的差的绝对值满足设定的最大允许误差值，则算法运行结束，此时群体中适应度值最大的个体，就是分配专家需求的最优解，即最合适的分配专家方案。

## 4.5 本章小结

本章首先结合学科平台系统的实际需求，描述和分析了专家分配问题，接着从专家属性的设置到约束条件的确定最后到目标函数的建立，层层深入地建立了分配专家问题的数学模型。然后，根据建立的数学模型，结合需求，在遗传算法的原理下，从编码方案的确定、生成初始化群体、适应度函数的设计、遗传进化、算法终止方面详细地描述了遗传算法在学科平台系统专家分配中的应用，最后，对遗传算法在学科平台系统专家分配中的应用流程结合流程图做了整体的介绍。

## 第5章 学科平台系统实现与遗传算法应用效果分析

前面的章节中已经对遗传算法的理论、系统的需求及设计、遗传算法在专家分配中的应用均做了详细地描述，本章将依据前面章节的设计方案来对学科平台系统中与分配专家有关模块的实现做详细介绍，并应用遗传算法进行分配专家，以及应用效果分析。

### 5.1 开发运行环境

学科平台系统开发运行环境如表 5.1 所示。

表 5.1 开发环境所需工具表

工具类型	工具名称及版本
操作系统	CentOS 7.2
数据库	Oracle 11g
开发语言	Java 1.8
Java 开发工具	IntelliJ IDEA 2018
版本控制	Git 2.10
类图设计	Visio 2013
软件包管理	Maven 3.6

### 5.2 系统的实现

本节主要是对学科平台系统中与专家分配相关的实现做详细描述，包括：材料管理模块的实现、分组管理模块的实现、专家信息管理模块的实现、专家遴选管理模块的实现、专家评议模块的实现。专家分配方法在学科平台系统的应用，保证了专家的科学性、合理性、公平性，很大程度上减轻了专家评议工作人员的工作量，间接地提高了专家评议质量。

#### 5.2.1 专家遴选的实现

专家遴选，也称为专家分配，是本系统的核心功能之一<sup>[57]</sup>。本系统分配专家方式灵活，提供手工分配专家和基于遗传算法的专家分配两种方式，当选用基于遗传算法分配



专家方式时,专家评议工作人员可以指定专家中的数量,并设置指定的参数和约束条件,由系统随机生成一套专家分配方案,完成初步分配专家过程。只有专家遴选管理员可以查看已生成的专家,根据学科平台的实际需求来选择是否启用该套专家。本节将展示基于遗传算法的专家分配页面、遗传算法中遗传算子代码和手工分配专家页面。

应用遗传算法进行分配专家时,只需要进入分组列表展示页面后,点击开始分配专家按钮就可以完成整个专家分配工作,页面如图 5.1 所示。



图 5.1 基于遗传算法的专家分配页面

当遗传算法运行结束时,就会得到最终的专家分配结果,即给出每个材料分组的专家编号等信息,如图 5.2 所示。

P1组中的专家 ✓								
编号	姓名	性别	所在单位	一级学科或专业学位代码	一级学科或专业学位名称	学科门类代码	学科门类名称	信用等级
204	张**	男	高校	1253	会计	12	管理学类	A
566	熊**	男	高校	202	应用经济学	2	经济学	A
599	丁**	女	高校	202	应用经济学	2	经济学	B
958	马**	男	高校	251	金融	2	经济学类	B
982	江**	女	高校	202	应用经济学	2	经济学	A
1084	卢**	男	高校	201	理论经济学	2	经济学	B
1169	余**	女	高校	202	应用经济学	2	经济学	B
1215	柴**	男	高校	201	理论经济学	2	经济学	A
P2组中的专家 ✓								
P3组中的专家 ✓								
P4组中的专家 ✓								
P5组中的专家 ✓								
P6组中的专家 ✓								

图 5.2 基于遗传算法的专家分配结果

本论文实现的遗传算法代码脱敏后托管于 Github 上,详见 5.3.2 节。下面将分别展示遗传算法的核心流程代码、选择算子、交叉算子、变异算子的代码。

1. 遗传算法的核心流程代码。遗传算法的核心流程是先计算并更新个体属性和种

群的属性，接着进行遗传进化，即进行选择、交叉、变异，最后根据进化代数和适应度要求条件判断是否结束运算，代码如图 5.3 所示。

```

299 //更新个体和种群属性，然后选择、交叉、变异。并循环此操作就是进化。每进化一代就需要重新计算个体和种群属性。
300 private void evolve() {
301     updateProperty();
302     //最佳保留和轮盘相结合。
303     optimumAndRouletteWheelSelection();
304     try {
305         averageCross();
306         singleMutate();
307     } catch (SQLException e) {
308         e.printStackTrace();
309     }
310 }

```

图 5.3 遗传算法核心流程代码

2. 选择算子。这里结合了轮盘赌模型和精英选择相结合的方法，能够保证分配的专家越来越接近目标，代码如图 5.4 所示。

```

274 //确定变异，当变异发生在最佳个体上时，需要判断变异是否变得更好。
275 private void mutate() throws SQLException {
276     Random random = new Random();
277     int[] expertGroupTmp;
278     for (int i = 0; i < populationSize; i++) {
279         double mutateRateTmp = random.nextDouble();
280         int pointTmp = random.nextInt(GeneticParameter.getGroupExpertSize());
281         if (mutateRateTmp <= GeneticParameter.getMutateRate()) {
282             expertGroupTmp = populationArray[i].randomGroup(populationArray[i].chrome.chromosome[pointTmp]);
283             if (i == bestIndividualIndex) {
284                 Individual tmp = populationArray[i].clone();
285                 populationArray[i].setGroupExpert(expertGroupTmp, pointTmp);
286                 populationArray[i].calculateProperty();
287                 if (populationArray[i].fitness < bestIndividual.fitness) {
288                     populationArray[i] = tmp.clone();
289                 } else {
290                     System.out.println(populationArray[i]);
291                 }
292             } else {
293                 populationArray[i].setGroupExpert(expertGroupTmp, pointTmp);

```

图 5.4 遗传算法选择算子代码

3. 交叉算子。代码中既实现了单点交叉方法，也实现了均匀交叉方法，在本文中使用了均匀交叉方法，如图 5.5 所示。

```

139 //单点交叉操作。
140 public void singleCross() {
141     int[] randomMatch = produceRandomMatch(populationSize);
142     Random random = new Random();
143     //确定交叉点
144     int point;
145     double crossRate;
146     for (int i = 0; i < randomMatch.length; i += 2) {
147         crossRate = random.nextDouble();
148         if (crossRate <= GeneticParameter.getCrossoverRate()) {
149             point = random.nextInt(GeneticParameter.groupExpertSize);
150             exchange(point, populationArray[randomMatch[i]], populationArray[randomMatch[i + 1]]);
151         }
152     }
153 }
154 //均匀交叉
155 private void averageCross() throws SQLException {
156     Random random = new Random();
157     int[] randomMatch = produceRandomMatch(populationSize);
158     for (int i = 0; i < randomMatch.length; i += 2) {
159         double averageCrossRate = random.nextDouble();

```

图 5.5 遗传算法交叉算子代码

4. 变异算子。这里实现了单点变异算子，如图 5.6 所示。

```

241 private void singleMutate() throws SQLException {
242     Random random = new Random();
243     for (int i = 0; i < populationSize; i++) {
244         double mutateRateTMP = random.nextDouble();
245         int groupIndex = random.nextInt(GeneticParameter.groupSize);
246         int expertIndex = random.nextInt(GeneticParameter.groupExpertSize);
247         int singleGene = produceSingleGene(populationArray[i].chrome.chromosome[groupIndex]);
248         if (mutateRateTMP <= GeneticParameter.SingleMutateRate) {
249             int tmp = populationArray[i].chrome.chromosome[groupIndex][expertIndex];
250             populationArray[i].calculateProperty();
251             double fitnessTMP = populationArray[i].fitness;
252             populationArray[i].chrome.chromosome[groupIndex][expertIndex] = singleGene;
253             populationArray[i].calculateProperty();
254             if (populationArray[i].fitness < fitnessTMP) {
255                 populationArray[i].chrome.chromosome[groupIndex][expertIndex] = tmp;
256             }
257         }
258     }
259 }

```

图 5.6 遗传算法变异算子代码

### 5.2.2 材料管理的实现

材料管理是专家评议的基础，由材料管理员将申请书材料信息录入系统，并对录入的材料进行基本的验证，并由教育厅对录入的材料进行资格审查，然后由材料分组管理员对材料进行分组，材料管理界面如图 5.7 所示。



图 5.7 材料管理界面

### 5.2.3 分组管理的实现

分组管理主要提供了定义分组、删除分组、修改分组、查询分组这四大基本功能，除此之外还提供了查询分组内的材料和专家信息的功能，如图 5.8 所示。



图 5.8 分组管理页面

### 5.2.4 专家信息管理的实现

专家信息管理主要提供了新增专家、删除专家、查询专家信息、修改专家信息，另外还提供了重置登录密码和修改登录密码的功能，如图 5.9 所示。

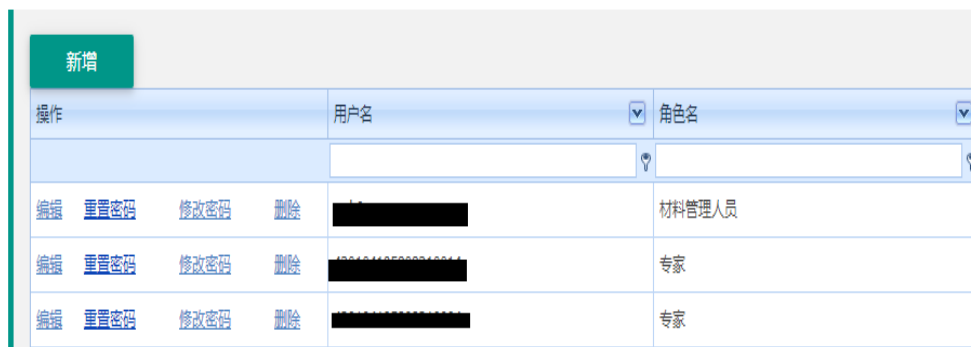


图 5.9 专家信息管理页面

### 5.2.5 专家评议的实现

专家登录系统后的首页上方有两个红色标注框，分别显示了该用户的身份和真实姓名，左侧菜单栏是该用户在本系统中具有的功能。如图 5.10 所示。

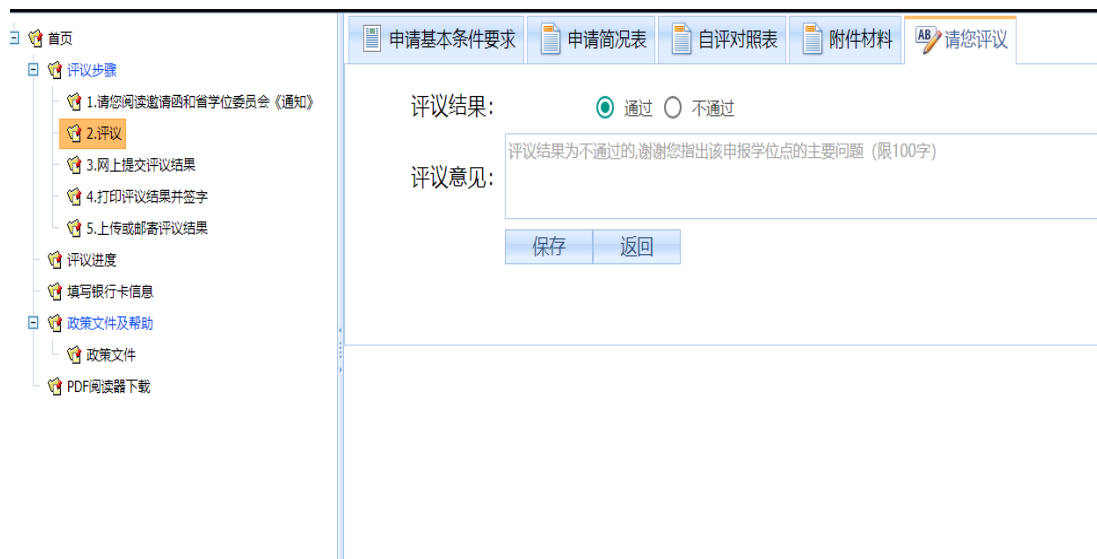


图 5.10 专家评议界面

## 5.3 应用效果分析

在本章上节中，已经对学科平台系统的主要功能实现模块进行了详细地描述，在本节中，将对遗传算法在学科平台系统专家分配方法中的实际应用进行分析，包括算法在

实际应用中参数的确定以及适应性分析。

### 5.3.1 实际效果测试结果

经过已有评审材料和专家库中的数据进行模拟分配试验，得出结论为：若系统对专家分配的时间要求不高，即不需要很强的时效性，那么随着遗传算法的不断进化，获得的专家分配结果也将越来越接近最优解。将遗传算法的参数设置为如 5.2 表所示。

表 5.2 实际权重参数设置表

权重	值	说明
$w^1$	0.6	专家与分组紧密度
$w^2$	0.1	专家数量
$w^3$	0.1	专家评议分组数
$w^4$	0.2	专家分配比例

实际遗传算法参数和其他参数如表 5.3 所示。

表 5.3 实际遗传算法参数设置表

参数名称	参数值
种群规模	60
交叉概率	0.6
变异概率	0.05
分组数目	10
每组分配专家数目	9
专家编号最小值	1
专家编号最大值	1300
专家最少评议分组数	1
专家最多评议分组数	1

根据上述参数设置，运行遗传算法后，所得的结果用折线图进行绘图，所得结果如图 5.11 所示。

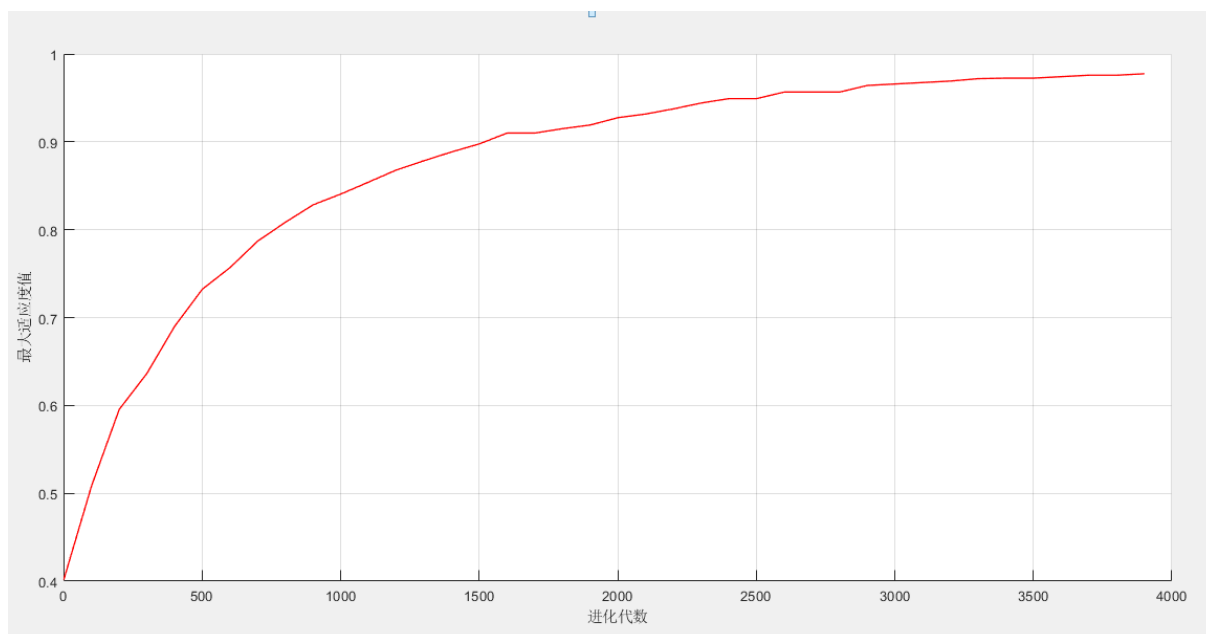


图 5.11 运行效果图

根据图 5.11 可知,随着进化代数的增加,种群中最大适应度值越来越大,表明专家分配得越好,即整个分配方案越佳。同时也可得知,在后期遗传算法进化得越来越缓慢,甚至出现进化停滞现象,而且进化代数越多,耗费的时间越长,因为遗传算法每进化一代,就需要算法运行一次。为了确定遗传算法在分配专家过程中实际应用时的运行参数,以及对遗传算法在实际应用的适用性分析,本节以系统试运行期间数据库中实际的材料为测试数据基准,对算法进行了相应的测试。

### 5.3.2 基于遗传算法的专家分配方法适用性分析

在上一小节中,遗传算法在学科平台系统分配专家方法中的具体运行参数,本文采用比较法进行分析,在同样的约束条件下,人工分配专家的方法和基于遗传算法的专家分配方法相比较<sup>[59]</sup>,基于遗传算法的分配方式在精度上和时间上明显优于人工分配专家的方式<sup>[60]</sup>。

试运行环境,一共有 298 份申请书材料,分成 20 组,每组分配 9 名专家,专家库中有 14000 名专家。采用人工分配专家方式,大约三天的时间,而基于遗传算法的分配方式大约需要 3 个小时,因此基于遗传算法的分配方式的效率明显占优势;人工分配专家方法在精度上约为 0.7 左右的匹配度,而基于遗传算法的分配方式在精度上约为 0.95 左右的匹配度,而且随着时间越长,遗传算法所得到的匹配度还会更高,基于遗传算法

的分配方式不容易犯错，分配专家更加合理。

由于学科平台系统具有一定的特殊性，在这种多重约束和庞大的专家的条件下，人工分配专家的方式是不合理的，也达不到专家评议工作人员专家评议的目的。在第二章中，已经详细描述了遗传算法的基本理论，由此可知，遗传算法在多约束条件下的全局搜索最优解的能力很好，因此本系统很适合采用基于遗传算法的专家分配方法来保证专家的质量，实现专家评议目的。本论文实现的遗传算法代码和测试数据脱敏后均托管于 Github 上，代码地址是：<https://github.com/igoodful/2019>。

### 5.3 本章小结

本章根据遗传算法的理论，实际的需求分析，以及遗传算法在学科平台系统专家分配方法中的应用设计，来描述系统的主要功能实现并对遗传算法的实际应用做了分析。首先，介绍了与专家分配有关的模块功能实现，通过图表以及部分核心代码的展示，使得实现过程的描述更加清晰，最后，根据学科平台的实际意义和目的，通过将遗传算法与人工分配专家模式进行对比，分析了遗传算法在学科平台系统中应用的适用性。



## 第6章 总结与展望

### 6.1 论文完成的主要工作

本文将遗传算法应用在学科平台系统中专家遴选中，结合实际工作需求，设计基于遗传算法的专家分配方法，减轻学科平台工作人员的工作量，从而提升效率。本文完成的具体工作如下：

1. 本文在遗传算法理论研究的基础上，充分理解遗传算法的基本构成要素和具体操作步骤及流程，并结合遗传算法特点，分析了算法在专家分配中的理论基础和可行性，并将遗传算法应用在学科平台系统中的专家分配模块中。

2. 本文研究了遗传算法在分配专家应用中的主要问题，对编码方案、适应度函数的设计、算法运行参数进行了研究分析，确定了在专家分配中应用遗传算法的流程。

3. 本文对学科平台系统的实际需求进行分析，并根据需求分析对系统的主要功能模块进行详细介绍。

4. 在遗传算法理论研究的基础上，结合实际需求，本文最终根据专家的一级学科或专业学位、信用等级、学科门类等属性，建立了多重目标优化下专家分配问题的数学模型。

5. 采用遗传算法实现专家分配，在分配专家的步骤上与基本遗传算法保持一致，但在某些步骤中根据实际的需求进行适当的增强，尤其在编码方案、初始化种群、适应度函数设计、遗传操作等方面做了深入研究。编码方案采用了实数编码，精简了约束条件，提高了效率；采用精英保留方法结合轮盘赌模型，既保证了种群的多样性，也防止了优良个体丢失，加快算法收敛速度；合理的适应度函数设计，可以适当的区分优劣个体，有效地解决了群体早熟问题采用终止进化代数和阈值结合使用来终止算法的运行，尽可能的在保证最优解的前提下，提升算法效率。

6. 根据需求分析，设计实现了学科平台系统。并借助实际应用场景，对遗传算法在学科平台系统实际应用的运行参数，结合经验，进行反复尝试，最终得出适合本系统的遗传算法运行参数。

## 6.2 未来工作展望

在完成以上六方面的研究内容的基础上,学科平台系统还有一些不足之处需要完善。后续主要的研究工作有:

1、虽然系统已经基本满足了当前学科平台的需求,但是,在给申请书材料分组上也可尝试应用遗传算法来实现自动分组功能,以减轻学科平台中的材料分组员的负担,同时也有利于后期的专家遴选。

2、目前的分配专家方法不可能十全十美,还需要不断地加强和改善,尤其是当一个分组内的材料混有不同的硕博类型、学科门类等因素,所以仍然要在分配专家中进行深入研究,力争让专家分配方法更完善、更高效。

3、目前对于分配专家方法中,尤其是遗传算法的相关技术应用中还有很多需要学习和研究的地方,在系统中还有很多地方可以完善的更好。因此,要不断地加强对理论知识的学习,并将理论与实际应用相结合,进一步提升自己的能力,争取在以后做得更好。

## 参考文献

- [1] 范军伟. 我国现行硕士、博士学位授权审核机制改革研究[D]. 兰州大学, 2010
- [2] 林梦泉, 龚桢楸. 浅谈我国学位授权审核的发展历程、作用与经验[J]. 学位与研究生教育, 2009(02): 6-11
- [3] 中国学位与研究生教育大事记(2017 年)[J]. 学位与研究生教育, 2018(05): 74-77
- [4] 教育部. 国务院学位委员会办公室负责人就《博士硕士学位授权审核办法》答记者问[J]. 新智慧, 2017, (06)
- [5] 刘洪名. 基于 C#的房屋中介管理信息系统的设计与实现[D]. 电子科技大学, 2012
- [6] 张素琪. 专家分配问题的启发式求解方法研究[D]. 河北工业大学, 2006
- [7] 李娜娜. 仿生算法及其在专家分配问题中的应用[D]. 天津大学, 2008
- [8] 宋存利. 求解混合流水车间调度的改进贪婪遗传算法[J/OL]. 系统工程与电子技术, 2019(03): 1-10
- [9] 朴雪涛. 大学制度创新与中国研究型大学建设[M]. 光明日报出版社, 2007
- [10] 张国宏, 周霞. 运用进化论完善我国科技管理同行评议专家制度[J]. 科研管理, 2015, 36(S1): 490-494
- [11] 张盼盼. 高校教学同行评议研究[D]. 山东师范大学, 2017
- [12] Liu R , Liu L . Predicting housing price in China based on long short-term memory incorporating modified genetic algorithm[J]. Soft Computing, 2019(11)
- [13] Mitchell M, Mitchell M. L.D. Davis, handbook of genetic algorithms[J]. Artificial Intelligence, 1998, 100(1): 325-330
- [14] 任子武, 伞冶. 自适应遗传算法的改进及其在系统辨识中应用研究[J]. 系统仿真学报, 2006, 18(1): 41-43
- [15] Maity S . An imprecise Multi-Objective Genetic Algorithm for uncertain Constrained Multi-Objective Solid Travelling Salesman Problem[J]. Expert Systems with Applications, 2015, 46: 1-18.
- [16] 陆杨. 专家选择算法的设计及系统实现[D]. 大连理工大学, 2005
- [17] 张华清. 柔性遗传算法用于阵列天线综合[A]. 中国电子学会. 2009 年全国天线年会论文集(上) [C]. 中国电子学会: 中国电子学会天线分会, 2009
- [18] 麻娟, 刘俨后, 楚满福, 高军. 求解第 II 类装配线平衡问题的混合遗传算法[J/OL]. 山东理

- 工 大学学报(自然科学版), 2019(03):55-59
- [19] 陈梦. 基于遗传算法的高校运动会系统的应用与研究[D]. 长江大学, 2018
- [20] 仁庆道尔吉. 车间作业调度问题的多目标模型建立及其算法[D]. 西安电子科技大学, 2013
- [21] 严丽军. 基于 GA 和 BP 算法的模糊神经控制在倒立摆中的应用[D]. 上海师范大学, 2005
- [22] 赵保才. 基于遗传算法的稳定 PID 设计[D]. 天津大学, 2006
- [23] 赵晓青, 戎晓剑. 一种自适应遗传算法对 Hopfield 网络权值的优化[J]. 石家庄铁路职业技术学院学报, 2007, 6(2): 84-87
- [24] 杜明芳. 基于改进遗传算法作业调度问题的优化及应用[D]. 西安理工大学, 2018
- [25] 杨林权. 机器人足球竞赛与设计技术[M]. 华中科技大学出版社, 2009
- [26] 沙宝良. 埋地管道的防腐参数采集传输与防腐性能评价[D]. 东北石油大学, 2013
- [27] 孙维亚. 稀疏分光 WDM 网络中基于遗传算法的组播业务研究[D]. 北京邮电大学, 2006
- [28] 谭阳. 求解广义旅行商问题的若干进化算法研究[D]. 华南理工大学, 2013
- [29] 王洁. 基于免疫遗传算法的 Job Shop 调度问题研究[D]. 浙江工业大学, 2009
- [30] 李静. 基于神经网络和遗传算法的 Ad hoc 网络故障管理模型研究[D]. 电子科技大学, 2008
- [31] 孙雁鸣. 基于遗传算法的无线传感器网络优化[D]. 辽宁工程技术大学, 2011
- [32] 李国, 郭海涛, 范大昭, et al. 一种基于遗传算法的高速影像匹配方法[C]. 系统仿真技术及其应用学术会, 2007
- [33] Jia Z , Gong L . Multi-criteria Human Resource Allocation for Optimization Problems Using Multi-objective Particle Swarm Optimization Algorithm[C] IEEE Computer Society, 2008
- [34] Long J , Szeto W Y , Huang H J . A bi-objective turning restriction design problem in urban road networks[J]. European Journal of Operational Research, 2014, 237(2): 426-439
- [35] 魏静. 并行遗传算法在生物序列比对中的应用研究[D]. 天津大学, 2004
- [36] 王景芳. 乙烯生产装置遗传网络模型的建立[J]. 化工自动化及仪表, 2005(05): 18-20
- [37] 客海生. 药品回收逆向物流的 VRP 模型研究[D]. 重庆交通大学, 2009
- [38] 董健康, 吴启明. 基于遗传算法的混合类型数据聚类分析[J]. 电脑知识与技术, 2009, 5(32): 8875-8876
- [39] 梁海葵. 糖厂蒸发工段的建模和控制[D]. 广西大学, 2008
- [40] 汤琳, 陶宏才. 基于相似度的交叉变异算子及其在分类规则挖掘中的应用[J]. 成都信息工程

- 学院学报, 2007(01): 64-68
- [41] 李良宝. 建设项目施工进度计划仿真研究[D]. 哈尔滨工业大学, 2007
- [42] 樊征兵, 宋亚辉, 张武林. 改进自适应遗传算法的平面传声器阵列优化[J]. 应用声学, 2017, 36(05): 429-437
- [43] 代冉. 免疫遗传算法及其在 TSP 中的应用研究[D]. 山东科技大学, 2008
- [44] 汪刚. 基于 java 的高职单招录取系统的设计与实现[D]. 苏州大学, 2010
- [45] Hu C , Ren G , Liu C , et al. A Spark-based genetic algorithm for sensor placement in large scale drinking water distribution systems[J]. Cluster Computing, 2017, 20(2):1089-1099
- [46] El-Mahdy O F M , Ahmed M E H, Metwalli S. Computer aided optimization of natural gas pipe networks using genetic algorithm[J]. Applied Soft Computing Journal, 2010, 10(4): 1141: 1150
- [47] 段会珍. 基于遗传算法的桁架结构可靠性优化设计[D]. 西安电子科技大学, 2012
- [48] 解庆, 赵小强. 遗传算法编码方法研究[J]. 甘肃科技, 2013, 29(02): 13-16
- [49] 李沛. 邮政金融 IT 项目的需求分析研究[D]. 北京邮电大学, 2010
- [50] 高林娥. 基因遗传算法的多目标优化问题的研究与应用[J]. 长沙航空职业技术学院学报, 2015, 15(2): 49-51
- [51] Afzalirad M, Shafipour M. Design of an efficient genetic algorithm for resource-constrained unrelated parallel machine scheduling problem with machine eligibility restrictions[J]. Journal of Intelligent Manufacturing, 2015: 1-15
- [52] Bradford E , Schweidtmann A M , Lapkin A . Correction to: Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm[J]. Journal of Global Optimization, 2018, 71(2)
- [53] 田志刚. 智能多目标优化理论及工程应用研究[D]. 大连理工大学, 2003
- [54] Ghiduk A S. Automatic generation of basis test paths using variable length genetic algorithm[J]. Information Processing Letters, 2014, 114(6): 304-316
- [55] Padillo F, José María Luna, Ventura S. A grammar-guided genetic programming algorithm for associative classification in Big Data[J]. Cognitive Computation, 2018
- [56] Saremi H Q , Abedin B , Kermani A M . Website structure improvement: Quadratic assignment problem approach and ant colony meta-heuristic technique[J]. Applied

- Mathematics & Computation, 2008, 195(1): 285-298
- [57] Salman A, Ahmad I, Al-Madani S . Particle swarm optimization for task assignment problem[J]. Microprocessors and Microsystems, 2002, 26(8): 363-371
- [58] Kolasa T, Krol D. A Survey of Algorithms for Paper-reviewer Assignment Problem[J]. IETE Technical Review, 2011, 28(2): 123-128
- [59] Wang F, Zhou S, Shi N. Group-to-group reviewer assignment problem[J]. Computers & Operations Research, 2013, 40(5): 1351-1362
- [60] Kale A, Kharat R, Bodkhe S, et al. Automated Fair Paper Reviewer Assignment for Conference Management System[C] Automation. IEEE, 2015

## 附录：攻读硕士学位期间发表论文及参与的项目

软件著作权

[1] \*\*与研究生教育管理信息系统 V1.0（软著），第一，2018SR589731，2018.7

[2] \*\*科学研究项目在线服务系统 V1.0（软著），第三，2018SR589735，2018.7

科研项目

[1] 项目名称：\*\*与研究生教育管理信息系统

项目时间：2017 年 6 月-2017 年 9 月

项目描述：通过该系统实现专家网上评议学科点，根据各个高校申报的学科点，然后根据相关政策和省学位办的要求进行申请书材料分组，然后根据要求给每组申请书材料分配指定数目的专家，当专家分配完成后，专家在指定的时间段内根据相关要求评议申请书材料，省学位办监督整个评议过程，确保评议顺利进行。

项目职责：主要负责的模块是申请书材料评议和后期的查询统计，参与了需求分析，Oracle 数据库开发及日常数据库管理，使用 PL/SQL 编写存储过程、函数、触发器。通过此项目熟悉了 oracle 数据库的日常操作与维护，更加熟悉了 Linux 系统，体会到了 C#开发的便捷性

开发平台及工具：Windows7、VS2013、SVN、Oracle 11gR2、PLSQL, C#、JAVA

[2]项目名称：\*\*乡舍

项目时间：2017 年 10 月-2018 年 6 月

项目描述：此项目采用了 MVC 设计模式，SSH 框架，Centos6.5 操作系统，mysql 5.7.22，Tomcat9.0，JAVA9.0 的技术上实现了\*\*乡舍项目。

项目职责：参与需求调研，需求分析，编写出实现房间详细搜索、农户请客申请、投资方请客申请等功能的完整代码，参与软甲开发和重大问题解决，数据库的备份与恢复，负责 Mysql 数据库的稳定运行，优化数据库性能。通过此项目熟悉了 SSH 框架和 MVC 模式，体会到了 java 的优美，体会到了 mysql 数据库高效轻便、操作简便的优点，同时也体会到了 Mysql 的调试困难等缺点。

开发平台及工具：Windows7、Idea、JAVA、SVN、Mysql5.7.22，VS2013