**分类号**  **学号**

**学校代码10487 密级**

****

**硕士学位论文**

**（学术型□ 专业型□）**

**标题：宋体，英文Times New Roman，一号，加粗，不超30字**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学位申请人** | **：** |  | **XXX** |  |  |
| **学科专业** | **：** |  | **XXXXX** |  |  |
| **指导教师** | **：** |  | **XXX 教授** |  |  |
| **答辩日期** | **：** |  | **XXXX年XX月XX日** |  |  |

**A Dissertation Submitted in Partial Fulfillment of the Requirements**

**for the Master Degree in Engineering/Science （工学/理学硕士）/the Professional Master Degree (专业学位)**

**英文题目，Times New Roman，小二号，  
实词的首字母大写**

**Candidate : LI Xinyao（中文习惯，姓在前且姓全部大写）**

**Major : \*\*\***

**Supervisor : Prof. XIAO Keling**

**Huazhong University of Science and Technology**

**Wuhan 430074, P. R. China**

**May, 2022**

**独创性声明**

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，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。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保 密□，在 年解密后适用本授权书。

本论文属于

不保密□。

（请在以上方框内打“√”）

学位论文作者签名： 指导教师签名：

日期： 年 月 日 日期： 年 月 日

# 摘 要

摘要是学位论文极为重要、不可缺少的组成部分，它是论文的窗口，并频繁用于国内外资料交流、情报检索、二次文献编辑等。其性质和要求如下：

1）摘要即摘录论文要点，是论文要点不加注释和评论的一篇完整的陈述性短文，具有很强的自含性和独立性，能独立使用和被引用。

2）硕士学位论文的摘要应包含全文的主要信息，并突出新见解与成果。

3）内容范围应包含以下基本要素：

（1）目的：研究、研制、调查等的前提、目的和任务以及所涉及的主题范围。

（2）方法：所用原理、理论、条件、对象、材料、工艺、手段、装备、程序等。

（3）结果：实验的、研究的、调查的、观察的结果、数据，被确定的关系，得到的效果、性能等。

（4）结论：结果的分析、研究、比较、评价、应用；提出的问题，今后的课题，建议，预测等。

（5）其他：不属于研究、研制、调查的主要目的，但就其见识和情报价值而言也是重要的信息。

4）摘要的详简度视论文的内容、性质而定，硕士学位论文摘要一般为500－600汉字。

5）摘要及全文中均建议不出现“我们”等字样。摘要中主语（作用）常常省略，因而一般使用被动语态；应使用正确的时态，并要注意主、谓语的一致，必要的冠词不能省略。

6）一般不用图、表、化学结构式、计算机程序，不用非公知公用的符号、术语和非法定的计量单位。

7）摘要中一般不使用缩写词，若实在需要，在第一次使用前，需给出中文全称（缩写词）；在使用英文缩写词之前，需给出英文全称（英文全称，缩写词），再次出现时可以采用中文或英文缩写词。

8）关键词应有3至8个，另起一行置于摘要下方，领域从大到小排列。关键词之间用分号隔开，最后一个关键词后面无标点。

9）摘要、关键词采用中文宋体；英文Times New Roman；小四号；1.5倍行距。

10）应有与中文摘要和关键词相对应的英文摘要和关键词。英语摘要用词应准确，使用本学科通用的词汇。

**关键词：** 关键词1；关键词2；关键词3

# Abstract

This is abstract.

英文摘要字体为Times New Roman，小四，1.5倍行距。

英文摘要和关键词应与中文相对应。英语摘要用词应准确，使用本学科通用的词汇；摘要中主语（作用）常常省略，因而一般使用被动语态；应使用正确的时态，并要注意主、谓语的一致，必要的冠词不能省略。

**Key words:** Keyword1, Key word2, Keyword3

# 目 录

[摘 要 I](#_Toc57978729)

[Abstract III](#_Toc57978730)

1 绪论（黑体，三号）

[1.1 研究背景与意义（黑体，四号） (2](#_Toc57978733))

[1.2 XXX国内外研究现状（请拟定具体的题目） (2](#_Toc57978734))

[1.3 存在的问题 (2](#_Toc57978735))

[1.4 本文主要内容 (2](#_Toc57978736))

2 系统与控制理论类

[2.1 引言（引言标题可选） (4](#_Toc57978738))

[2.2 预备知识（可选，标题可自选） (4](#_Toc57978739))

[2.3 问题的描述（请拟定具体的题目） (5](#_Toc57978740))

[2.4 控制器设计与闭环系统分析（请根据所设计的控制器特点自行拟定具体的题目） (6](#_Toc57978741))

[2.5 数值仿真（请拟定具体的题目） (6](#_Toc57978742))

[2.6 本章小结 (6](#_Toc57978743))

3 理论/算法类研究类论文

[3.1 引言（引言标题可选） (7](#_Toc57978745))

[3.2 \*\*理论/算法 (7](#_Toc57978746))

[3.3 \*\*仿真或算法实现 (7](#_Toc57978747))

[3.4 理论/算法准确性的评估 (7](#_Toc57978748))

[3.5 分析与讨论 (7](#_Toc57978749))

[3.6 本章小结 (7](#_Toc57978750))

4 学位论文写作细则

[4.1 关于图 (8](#_Toc57978752))

[4.2 关于表格 (9](#_Toc57978753))

[4.3 名词、术语 (10](#_Toc57978754))

[4.4 符号、单位的使用 (11](#_Toc57978755))

[4.5 数字的使用 (11](#_Toc57978756))

[4.6 其它应该注意的问题 (12](#_Toc57978757))

[4.7 本章小结 (13](#_Toc57978758))

[5 总结与展望](#_Toc57978759)

[5.1 本文主要内容及结论 (14](#_Toc57978760))

[5.2 本文的主要创新点 (14](#_Toc57978761))

[5.3 展望 (14](#_Toc57978762))

[致 谢 (15](#_Toc57978763))

[参考文献 (16](#_Toc57978764))

[附录1 攻读硕士学位期间取得的研究成果 (19](#_Toc57978765))

[附录2 其它附录 (21](#_Toc57978766))

目录编写要求：

1. 目录自动生成时，只需列出二级标题；
2. 中文宋体，英文和数字Times New Roman，四号，行距：固定值25磅；
3. 注意避免出现不具辨识度的标题：如材料与方法、实验结果、分析与讨论、结论，若属于这种类型的论文时，可以把结果部分拆开，根据结果拟定题目。

# 绪论

硕士学位论文应能表明作者确已在本门学科上掌握了坚实的基础理论和系统的专门知识，并对所研究课题有新的见解，有从事科学研究工作或独立担负专门技术工作的能力。硕士学位论文的字数一般不少于2.5万字，加上各类图表，从绪论到总结与展望的总页数一般不少于50页。否则，可能显得工作量不够饱满。

学位论文题名是以最恰当、最简明的词语反映论文中最重要的特定内容的逻辑组合。题名既要准确地描述内容，又要尽可能地精练，一般不宜超过30个字。题名一般避免使用不常见的缩略词、字符、代号和公式等。若实属必要，需要摘要中给出中英文全称与缩写。学位论文内容应结构合理、立论正确、推理严谨、文字简练、层次分明、逻辑严密、数据真实可靠。

学位论文文字排版的字号、行距、字距的大小，以版面清晰、容易辨识和阅读为原则。为统一起见，具体要求如下：

（1）论文页眉，楷体，字体16.5；

（2）章和节的题名用黑体，字号分别用三号和四号；

（3）正文内容用宋体，英文用Times New Roman，小四号；行间距1.5倍；正文注意两侧对齐。

（4）除了章的标题使用1、2、3、….进行编号外，正文中的其它编号不能再使用诸如1、2、3、……形式的编号（该编号已用于章的编号），只能使用诸如（1）、（2）、（3），……，或1）、2）、3），……，或①、②、③，......等形式的编号。

绪论部分是整篇论文的导引，应包括选题背景、意义；国内外研究概况；前人研究中存在的问题或知识空白；进而引出本文的研究设想，简要给出全文各章节的主要内容、以及章节之间相互联系。

在写作中无论是研究背景及意义，还是国内外研究现状，要做到有依据都必须引用参考文献。通常情况下，绪论部分的参考文献引用应占全文参考文献引用的80%以上。参考文献引用的顺序必须是按照在文章中出现的先后顺序进行排列。

以下简要说明一下绪论部分的内容及各级标题格式等。

## 研究背景与意义

大数据时代随着互联网持续发展而到来，各行各业的数据量积累急剧增长，大数据推动各领域进步成为研究热点之一[1]。Web、移动端、智能穿戴、监控设备等每天都产生大量的数据。计算机视觉、自然语言处理等人工智能领域的发展促进了大量高纬度数据的产生。如何高效处理海量数据，利用大数据获得有价值的信息成为各个企业、科研机构重点关注的领域。

数据挖掘技术随着大数据时代的到来被广泛研究并取得很大成功[2]。数据挖掘是可以从大量数据中提取有效信息的技术。数据挖掘依靠运筹学、统计学、人工智能等技术，高效地分析数据，从数据中挖掘出潜在的、事先不知道的、潜在有用的信息。数据挖掘在很多另有都用广泛应用，尤其是教育、风控、医疗、银行、电信等领域。

聚类是数据挖掘中的主要技术之一。聚类可以将指定数据集按照相似度分为特定的几类，使得同一个簇内数据的相似度尽可能大，而被分配在不同簇内的数据的相似度尽可能小，进而对有针对性地对不同簇开展研究[3][4]。聚类是一种无监督的学习方法，区别于有监督学习需要提供已知的分类信息来对分类器进行训练，聚类算法仅依靠数据间的相似程度。

随着聚类在各个领域的应用，聚类的平衡性体现出了重要的应用价值。分布式系统中，可以用聚类来决定集群中机器负责的数据块，聚类的应用使得具有更高相似性的数据处于同一计算节点，分布式系统间的通信消耗得到下降。然而，聚类算法没有对不同簇中节点数量做出约束，导致不同簇中点数偏差可能过大，某些簇中数据块可能过多或者过少。簇中数据节点数量分布不平衡会导致分布式系统的性能降低。在分布式系统中，聚类的平衡性决定了资源分配的均衡性，决定了系统的整体性能。在图像分类领域中，词袋模型（bag-of-words）需要K-means算法对图像的特征进行聚类，将图像分为不同的类别。文献[5][6]指出，聚类的平衡性决定了图像分类的准确性，而K-means算法不能保证聚类的平衡性。在MBA学生分组问题中[7]，作者考虑了MBA团队的平衡约束。

由上所述，在聚类中如何保证不同簇间的平衡性有重要的应用价值[8]。在聚类迅速发展的几十年中，大多数学者致力于聚类算法准确性的研究中，在算法平衡性的关注上相对较少。本课题针对平衡性聚类问题进行研究，在满足不同簇之间平衡性的前提下，持续算法准确性。

## 平衡聚类问题国内外研究现状

本节将从三个方面进行介绍。首先介绍聚类问题及其相关变种问题，然后介绍聚类问题的研究现状，最后介绍平衡聚类问题的研究现状。

### 聚类问题

聚类旨在通过数据内在特性对数据进行分类，是数据挖掘中一种重要的技术手段。最早的聚类算法在上世纪六十年代被提出。主要的聚类算法可以分为：划分聚类、密度聚类、网格聚类、层次聚类[9]。聚类算法的分类如图2.1所示。

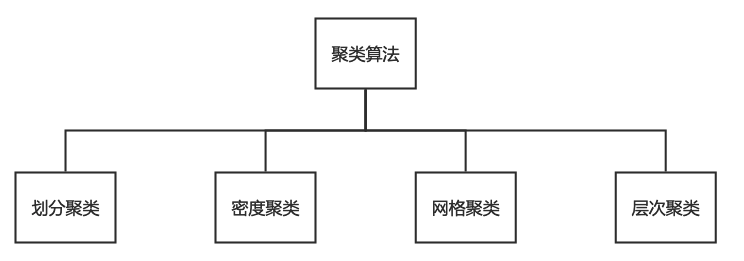


图4.1 G11的进化情况

#### 划分聚类

划分聚类首先确定划分组的数量K，然后不断在组之间进行对象的重新分配，最终达到收敛。划分聚类中比较有代表性的是k-means算法。在k-means中，会将每个点分配给距离质心最近的那个簇。其中质心是一个簇中所有点的平均值，质心的坐标是簇中所有点的平均值。K-means实现起来简单，时间复杂度是O（nkt），其中n是样本数量，k是簇的数目，t是算法迭代次数，算法时间复杂度近于线性，因此在聚类问题上经常使用。K-means的伪代码流程如下：

1. 从数据点中，选择K个簇中心；
2. 样本点分配到距离最近的簇中心；
3. 计算簇中点的均值，并作为新的簇中心；
4. 如果有簇中心被重新指定，继续执行步骤2，否则结束。

通过伪代码我们可以发现，k-means在步骤（1）中的簇中心选择是随机的，不同的处始状态会得到不同的聚类结果，算法对初始状态的依赖较高。减少k-means算法对初始解的依赖性是改善k-means算法的重要手段。在文献[10]中提出了k-means++的方法，针对初始中心的生成进行优化，降低了处于边缘地带的点称为簇中心点概率。

#### 密度聚类

如果簇的形状不是规则或者相互交错，基于距离的聚类算法的表现往往会变差。在这种情况下，基于密度的聚类算法表现更加优秀。基于密度的聚类可以发现任意形状的聚类，并且对噪声敏感度较低。密度聚类中比较经典的是DBSCAN[11]首先设置一个数量阈值，根据样本附近的点数量，分为低密度和高密度区域。DBSCAN是经典的密度聚类之一，该算法需要输入半径Eps和阈值MinPts。以指定点为圆心，Eps为半径，如果圆中点的数量大于等于MinPts，则将圆心称为高密度点；否则，将圆心称为低密度点。如果两个高密度点的距离小于半径，则将两个点放到一个簇中；如果高密度点和低密度点间距离小于半径，则将低密度点加入高密度点的簇中，并称低密度点为边界点。通过这种方式生成簇不需要指定簇数量k。DBSCAN算法的缺点是优度依赖于半径Eps和阈值MinPts，确定这两个参数的最优值是一个新的挑战。

#### 网格聚类

网格聚类将聚类空间转换为网格结构，并在这个网格结构上进行聚类操作。典型的网格聚类代表是CLIQUE[12]。CLIQUE是一种同时基于网格和密度的聚类方法，在网格结构中进行密度聚类。

#### 层次聚类

层次聚类主要分为两种：（1）自底向上地将不同簇中的样本合并，直到满足停止条件；（2）认为所有点在一个簇中，对簇进行拆分，直到满足停止条件。其中DIANA[13]是一种典型的自底向上的合并聚类方法。CURE[14]是一种典型的自顶向下的方法。

### 平衡聚类问题

BMSSC在传统聚类问题的基础上加入了平衡约束，任意两个簇中点的数量差距不能超过1。加入约束后平衡聚类仍然是np-hard问题。平衡聚类问题的应用非常广发，但是这方面的研究却相对较少。平衡聚类问题首次被研究是2014年，Malinen[15]等人在k-means算法上加入平衡约束来解决BMSSC问题。此算法能够解决超过5000个数据点的聚类问题。Costa[16]等人在2016年提出了一种变邻域算法来求解BMSSC问题，此算法优于当时其他算法求得的结果。2021年，Qing Zhou[17]等人将进化算法、阈值搜索等启发式算法应用到BMSSC问题上，在算法结果、稳定性上都优于2016年Costa提出的变临域搜索算法，成为当前解决BMSSC最好的算法。

## 存在的问题

综合当前国内外研究现状，总结现有研究中存在的问题，以便引出本文的研究内容。

在绪论中提出问题，论文的工作正是围绕这些问题而展开。

## 本文主要内容

不同学生学位论文的类别各异，大致可分为工程类，理论类、算法研究类、仪器/工艺设计与研发类、综合类等。不同论文的章节结构各异，但每种类型的论文还是有其特定的格式。原则上，硕士论文的主体研究内容不得少于2章，加上绪论 、总结与展望，累计不得少于5章。

作为华中科技大学硕士学位论文模板，本文首先给出了不同类型的研究论文典型结构供大家参考，再根据学术出版的规范化要求，说明论文写作中的细则。全文共分为5章。主要内容如下：

第一章 绪论：简要介绍论文的研究背景、国内外研究现状、存在的问题，给出全文的主要研究内容。为了让读者更容易理解全文，建议用一个文档结构图给出各章节逻辑关系。

本文共分为5章内容，章节内容之间的关系图如图1.1所示。

图1.1 组织结构图

第二章 系统与控制理论类论文结构：……

第三章 算法研究类结构：……

第四章 学位论文写作细则：……

第五章 总结与展望：给出全文的主要内容及结论，总结本文的创新点，并对未来的工作进行展望。

# 平衡聚类问题

平衡聚类（BMSSC）问题是一个具有NP难度的组合优化问题。随着聚类问题的应用，聚类问题平衡性得到广泛关注。平衡聚类问题在分布式系统、图像处理、MBA学生分组等问题上都有实际应用。传统的聚类算法如K-means、DBSCAN无法保证聚类结果的平衡性。本章首先对平衡聚类（BMSSC）问题进行描述，其次建立了BMSSC问题的数学模型，最后通过分析平衡聚类问题的特点，引出了一种基于阈值搜索的启发式算法。

## 平衡聚类问题描述

在欧几里得空间中给定一个包含个点的数据集，将数据集中的数据尽可能平均地分配到个簇中，在相同簇中的数据点关联程度较高，不同簇中的数据点之间关联程度较低。聚类问题中的关联程度通常由欧几里得距离的平方进行表示：，在使用欧几里得距离平方表示关联程度的时，聚类问题也可以称为最小平方和聚类问题（minimun sum-of-squares clustering，MSSC）。聚类问题的目标是要找到个簇，并且使得每个点到其所在的簇中心的欧几里得距离平方总和最小。平衡聚类问题在聚类问题的基础上加入了平衡性的约束，即任何两个簇中数据数量差距不能超过1。通过对问题分析可知，平衡聚类分配结果必然包含 个包含个数据点的簇，和个包含个数据点的簇。对于相同算例数据，传统聚类问题的划分结果如图2.1所示，平衡聚类问题的划分结果如图2.2所示。

BMSSC是np难问题。在加入平衡性约束后，当的时候，BMSSC问题可以等价于最小加权完全匹配问题（balanced minium sum of squares clustering，BMSSC），此问题可以在时间复杂度内得到解决。但是对于的情况，BMSSC仍然是难问题，很难找到多项式时间复杂度的算法。

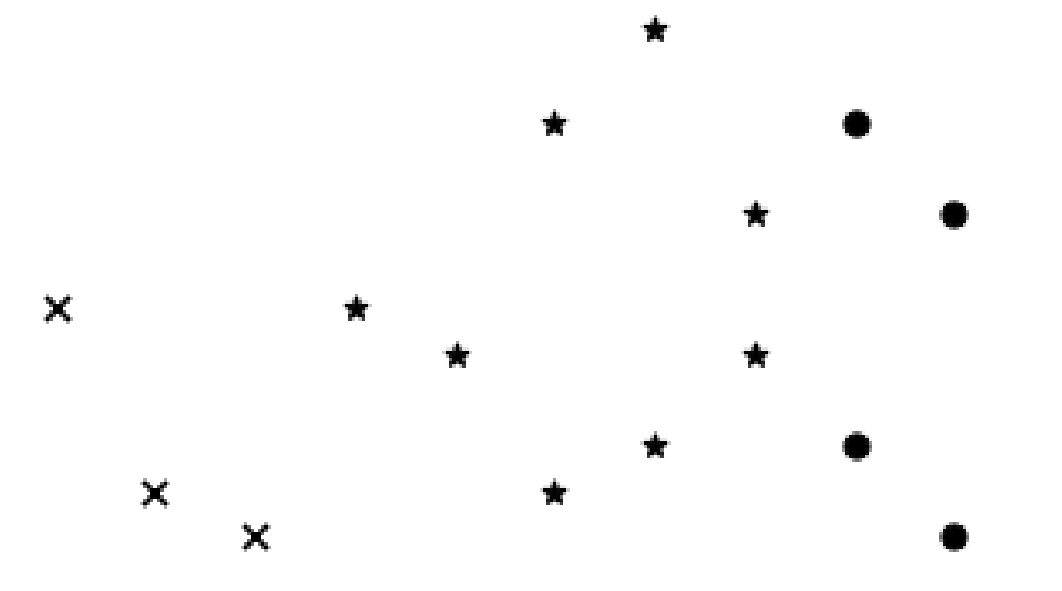


图2.1 聚类划分结果

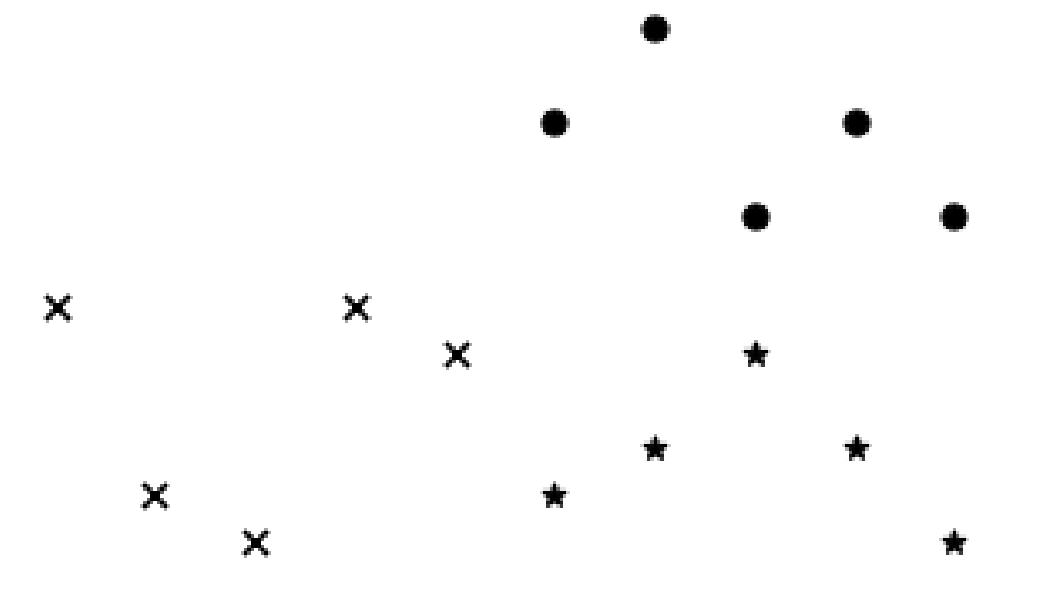


图2.2 平衡聚类划分结果

## 数学模型

给定一组n个数据点，在一个d维欧几里得空间中，其中数据点的每个维度的值是可以是正的、负的或0的实数，MSSC是对给定的相交的子集，使得每个数据点到其群集质心的平方距离之和被最小化。这个问题可以用如下的公式进行描述：

上述公式中，表示数据点被分配给簇，表示数据点没有分配给簇，第二个公式约束了点只能被分配到一个簇内。

其中是簇的质心，，每一维的坐标值都是此簇中所有点的算数平均值。表示两个点之间的欧几里得距离，可以表示为 。

BMSSC相比于MSSC多了一个任意两个簇中点的差距不能超过1的约束，BMSSC的目标不仅要最小化每个样本到其簇中心的总和，还要保证任意两个簇都是平衡的，这个约束可以用以下公式来表示。

当的时候，任意两个簇之间的点数绝对相等；当的时候，有个簇中点的数量分别比其余每个簇的点数多1。

## 问题分析和算法设计思路

本节首先对最小平方和平衡聚类问题（BMSSC）的特点进行分析，其次针对BMSSC的问题特点获得算法设计思路。

### 问题特点分析

（1）数据点在不同簇中的分配会相互影响。对于最小平方和平衡聚类问题（BMSSC），每个数据点在分配阶段要考虑平衡性约束，即当前本簇数据点数量不能超过本簇最大容量，这种平衡性约束给求解BMSSC问题增加了难度。而且，每个簇的中心都是通过计算本簇数据点平均值得到的，改变簇中数据点会直接改变簇中心，这给求解BMSSC问题带来了阻碍。

（2）BMSSC问题是具有NP难度的大规模组合优化问题，问题的求解非常耗时。BMSSC问题是最小平方和聚类问题（MSSC）的变种，在分布式系统数据分布，MBA学生分组等问题上均有应用。随着BMSSC问题规模逐渐增大，求解问题所需的时间和存储空间会越来越大。在这种情况下，使用精确算法（Exact Algorithm，EA）获得最优解变得十分困难。在实际应用场景中，往往不需要取得全局最优解，可以接受计算成本更低的较优解。

### 算法设计思路

求解组合优化问题的解决方法主要分为两类：精确算法和启发式算法。提出的问题的解决方法主要分为两类，精确算法和启发式算法。由于启发式算法效率更高，在相同时间内获得解的质量更高，本课题主要探讨的是启发式算法。启发式算法是一种基于经验或者直观构造的算法，该算法能在给定时间内求解出组合优化问题的较优解。局部搜索算法是用来求解组合优化问题的一种经典的启发式算法。局部搜索最终解质量与初始解、邻域结构、扰动策略的设计息息相关。优秀的初始解为局部搜索算法提供合理的方向，减少局部搜索消耗的时间。合理的邻域结构令局部搜索探索更有效的空间。扰动策略可以使局部搜索算法避免陷入局部最优。因此，有效的初始解算法、邻域结构、扰动策略可以令局部搜索算法兼具集中性和发散性，进而找到更优秀的较优解。混合进化算法往往在求解组合优化问题上有着出色的效果。混合进化算法通常由种群、交叉算子、局部搜索构成。交叉算子用于结合两个父代的特点，生成新的子代。混合进化算法通过将交叉算子与局部搜素相结合保证了种群的优度和多样性。

为了提升BMSSC问题的求解效率和求解质量。基于BMSSC的问题特点，为局部搜索算法设计了高效的邻域结构及扰动策略。本课题根据初始解对局部搜索算法的影响在初始解构造阶段提出了一种基于随机的贪心算法；基于xxx这一特点，设计了一种xxx的邻域结构；根据xxx这一特点，设计了xxx的扰动策略。

综上所述，为了更高效求解BMSSC问题，本课题提出了一种混合进化算法。该算法主要分为两个阶段，第一阶段通过带有随机性的贪心算法构造合法的初始解，第二阶段使用混合进化算法对解进行迭代改进。

## 本章小结

本章首先对平衡最小平方和聚类（BMSSC）问题进行描述，并构建了BMSSC问题的数学模型，主要包括该问题的已知条件、决策变量、约束及目标函数。然后分析了BMSSC问题的特点：该问题是NP难问题，很难在可接受的时间范围内找到最优解；数据点在不同簇中的分配会互相影响。最后根据BMSSC问题的特点提出了一种基于带有随机性的贪心算法和混合进化算法，同时简单介绍了该算法的设计思路。

# 基于随机初始化和混合进化的启发式算法

第二章对BMSSC问题进行了描述，对该问题进行数学建模并分析了该问题特点，引出了一种基于随机出初始化和混合进化（RD-HEA）的启发式算法来求解BMSSC问题。本章详细介绍了RD-HEA算法。首先，种群中的个体使用随机贪心策略生成初始解，为种群提供多样且有潜力的优化方向。然后基于混合进化算法，对种群进行随机交叉并生成新解。同时引入响应阈值搜索进行扰动，对生成的新解进行持续改进。

## 算法总体结构（贪心初始化后需要局部搜索么）

RD-HEA是一个两阶段算法：第一阶段通过带有随机性的贪心算法对种群中个体进行初始化，这些个体能够为后续的混合进化算法提供一个优秀的探索方向；第二阶段通过混合进化算法生成子代，通过执行响应阈值搜索算法。算法中设计了xxx的邻域结构和邻域动作选择策略，使得局部搜索过程能够快速收敛。同时使用了基于阈值的随机扰动、基于xx的随机扰动来增加局部搜索的发散性。

图3.1以流程图的方式展现了RD-HEA的算法框架。从图中可以看出初始化种群阶段会对种群中所有个体应用带有随机性的贪心初始化算法。带有随机性的贪心初始化算法可以保证种群中个体的优度和多样性。之后进入混合进化阶段，混合进化是一个循环迭代的过程，每次迭代在种群中随机选择两个解作为父类，对两个父类应用交叉算子生成子类。之后对子类应用响应阈值搜索，对子类解质量进行提升。最终，得到优秀的子类解，并将子类解与种群中各个解进行对比，应用种群更新策略对种群中解进行更新。

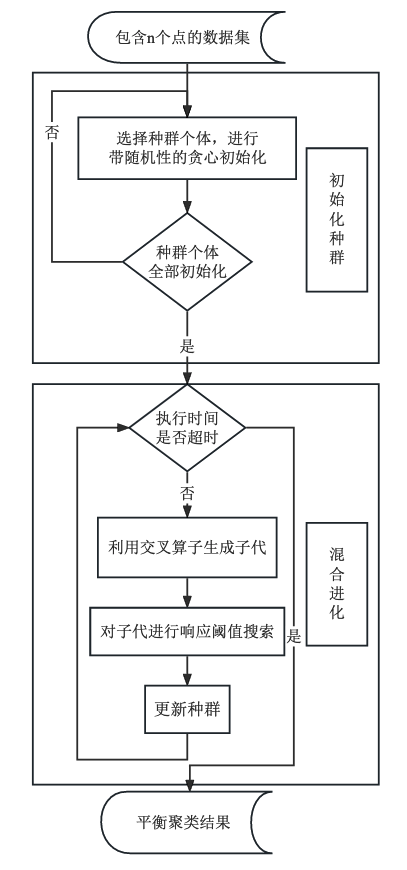


图3.1 RD-HEA算法总体框架

## 初始解构造

初始解是局部搜索的起始位置，初始解的质量也决定着局部搜索最终解的质量和局部搜索所消耗的时间。拥有质量越好的初始解，进化算法和局部搜索算法就有更大的概率向最优解的方向进行优化。因此，如何利用算法生成一个好的初始解是RD-HEA算法关注的重点问题之一。RD-HEA的初始化算法有两种方式：随机初始化算法和带有随机性的贪心初始化算法。本文种群部分个体采用随机初始化算法，剩余个体采用带有随机性的贪心初始化算法，增加了个体间的多样性和差异性。

### 带有随机性的贪心初始化算法

算法3.1用伪代码的形势描述了RD-HEA的带有随机性的贪心初始化算法。其中输入是包含N个数据点的集合，集合中的每个点都是在d维欧几里得空间中，以及数据点要被分成的簇数。算法的第1行和第2行对用到的中间变量进行赋值，代表当前进行分配的簇，代表剩余待分配的数据点集合。第三行对初始解构造算法的簇分配结果进行初始化，是一个长度为的数组，数组的值是数据点被分配到的簇。第5-9行对应的是确定每个簇中元素的数量。第10行从剩余待分配的数据点集合中随机选择一个节点作为簇的中心点。第11行在剩余待分配数据点集合中选择个距离中心节点最近的数据点（包含中心节点本身）。第12-15行将选择出来的个数据点分配给当前簇，并从剩余待分配集合中去除这个数据点。

|  |  |
| --- | --- |
| 算法3.1：RD-HEA带有随机性的贪心初始化算法（GreedyInitialSolution） | |
| 输入：个数据点的集合，要分成的簇数 | |
| 输出：BMSSC问题的初始解 | |
| 1. |  |
| 2. |  |
| 3. |  |
| 4. | while do |
| 5. | if then |
| 6. |  |
| 7. | else |
| 8. |  |
| 9. | end if |
| 10. | /\*随机选择数据点\*/ |
| 11. | /\*选择最近数据点集\*/ |
| 12. | for do |
| 13. |  |
| 14. | end for |
| 15. | /\*更新未使用的点\*/ |
| 16. | end while |
| 17. | return |

### 随机初始化算法

|  |  |
| --- | --- |
| 算法3.1：RD-HEA随机初始化算法（RadomInitialPopulation） | |
| 输入：个数据点的集合，要分成的簇数 | |
| 输出：BMSSC问题的初始解 | |
| 1. |  |
| 2. |  |
| 3. |  |
| 4. | while do |
| 5. | if then |
| 6. |  |
| 7. | else |
| 8. |  |
| 9. | end if |
| 10. | /\*随机选择数据点\*/ |
| 11. | /\*选择最近数据点集\*/ |
| 12. | for do |
| 13. |  |
| 14. | end for |
| 15. | /\*更新未使用的点\*/ |
| 16. | end while |
| 17. | return |

## xx局部搜索策略

## 本章小结

# 实验结果

学位论文很多的错误源于凌乱的格式。

为规范学位论文写作，本章结合工科学位论文的特点，参照学术出版规范，就图、表制作，名词、术语、单位、符号、数量等的使用规范化进行了说明。

## 测试算例

## 测试环境测试

## 实验结果对比

## 算法关键策略分析

## 本章小结

# 总结与展望

## 全文总结

对全文进行全面地总结，并根据各章节归纳出若干有机联系的论点。按正文的内容分段描述，包括本研究“做了什么（提出\*\*新理论/算法、设计或研发\*\*工艺/仪器）、获取什么结果、得出什么结论”。

请特别注意，全文总结与摘要及各章的小节要有所区分，不能简单的拷贝。这里的重点是结论，结论应该准确、完整、明确、精练。

## 本文的主要创新点

通常情况下，学位论文的创新点应放在最后一章。

创新点要凝炼，表述要清晰明了，如提出了什么创新的思路，主要特点是什么，相比现有理论或技术的提高是什么、或者有什么新的发现，是否具有重要的科学意义或应用前景。既不能过于简单，也不要太细。

硕士学位论文创新点不宜太多，一般为2个左右即可，要注意归纳创新点，千万不要以为越多越好。论文的创新不以创新点的多少来评定的，而以其创新性的价值来评定。几章的工作合在一起凝炼成一个创新点也不是不可以的。

## 未来工作展望

对本研究成果的意义、推广应用的现实性或可能性加以论述。同时，描述本文研究中尚存在的不足，或因时间尚未完成但又必须继续的工作，对进一步的工作进行展望。

# 致 谢

对在课题研究及论文写作过程中给予指导和帮助的导师、校内外专家、实验技术人员、同学等表示感谢。

在致谢时建议具体，不同的人如何助力完成你的论文，都需要特别注明。如导师、其他老师或实验技术人员、以及同学对你论文的贡献是不一样的，有指引课题方向、修改论文，也有具体教会实验操作，也有协助你做了哪方向的实验，或者给你精神安慰、陪你度过紧张的研究生生涯。

越具体越能表达你真实的感受，否则就是毫无意义的套话。

注意：部分论文存在导师与致谢中主要表示感谢的老师不一致的情况，甚至出现致谢中提及的老师、发表论文的作者、申请专利的发明人等都与封面上的导师无关的情况，应避免出现这种情况，可以将两位老师均作为论文指导老师。

# 参考文献（格式不合格）

[1] Rabhi L , Falih N , Afraites L , et al. Big Data Approach and its applications in Various Fields: Review[J]. Procedia Computer Science, 2019, 155:599-605.

[2]袁红春, 熊范纶, 淮晓永. 空间数据挖掘及其与智能系统的集成框架[J]. 信息与控制, 2002, 31(4):6.

[3] Ghosal A , Nandy A , Das A K , et al. A Short Review on Different Clustering Techniques and Their Applications[J]. 2020.

[4] Higuchi A , Maehara R . A factor-cluster analysis profile of consumers[J]. Journal of Business Research, 2021, 123.

[5] Wu J , Rehg J M . Beyond the Euclidean distance: Creating effective visual codebooks using the Histogram Intersection Kernel[C]// 2009 IEEE 12th International Conference on Computer Vision (ICCV). IEEE, 2009.

[6] Wu J , Tan W C , Rehg J M . Efficient and Effective Visual Codebook Generation Using Additive Kernels[J]. Journal of Machine Learning Research, 2011, 12(7):3097-3118.

[7] Desrosiers J, Mladenović N, Villeneuve D. Design of balanced MBA student teams[J]. Journal of the Operational Research Society, 2005, 56(1): 60-66.

[8]唐海波, 林煜明, 李优. 一种基于K-Means的平衡约束聚类算法[J]. 华东师范大学学报（自然科学版）, 2018, 000(005):164-171.

[9] Madhulatha T S . An Overview on Clustering Methods[J]. IOSR Journal of Engineering, 2012, 2(4).

[10] Arthur D , Vassilvitskii S . K-Means++: The Advantages of Careful Seeding[C]// Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007. ACM, 2007.

[11] Ester M . A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise[J]. Proc.int.conf.knowledg Discovery & Data Mining, 1996.

[12] Agrawal R , Gehrke J E , Gunopulos D , et al. Automatic Subspace Clustering of High Dimensional Data for data mining applications[J]. Data Mining and Knowledge Discovery, 1998, 27(2):94-105.

[13] Kaufman L , Rousseeuw P J . Finding Groups in Data: An Introduction to Cluster Analysis[M]. DBLP, 1990.

[14] Guha S , Rastogi R , Shim K . CURE : An Efficient Clustering Algorithm for Large Databases[J]. Information Systems, 1998, 26(1):35-58.

[15] Fränti, P., Brown, G., Loog, M. , et al. Balanced K-Means for Clustering[J]. 2014, 10.1007/978-3-662-44415-3(Chapter 4):32-41.

[16] Leandro R. Costa, Daniel Aloise, Nenad Mladenović. Less is more: basic variable neighborhood search heuristic for balanced minimum sum-of-squares clustering[J]. Information ences, 2017, 415:247-253.

[17] Zhou Q , Hao J K , Wu Q . Responsive threshold search based memetic algorithm for balanced minimum sum-of-squares clustering[J]. Information Sciences, 2021, 569(4).

[18] Bertoni A , Goldwurm M , Lin J , et al. Size Constrained Distance Clustering: Separation Properties and Some Complexity Results[J]. Fundamenta Informaticae, 2012, 115(1):p.125-139.

# 附录1 攻读硕士学位期间取得的研究成果

**发表与接收论文**

[1] 参照参考文献列出学术论文相关信息（含期刊、会议、或参编书稿），但无论有多少个作者，都必须列出全部作者名；若为英文论文，则名在前姓在后，姓名均为全称；在本人的名字加粗，以示区别（若为第一作者，则需在最后特别注明署名华中科技大学是否为第一单位）

[2] 若已发表，按参考文献给出页码；若只是online,给出链接；若接受或修改或投稿或拟投，也必须分别注明

[3] 一般情况，一作或重要的论文放在前面

**发表论文（举例）**

[1] Hajiali, Faezeh, **Saeid Tajbakhsh**, and Akbar Shojaei. Fabrication and properties of polycaprolactone composites containing calcium phosphate-based ceramics and bioactive glasses in bone tissue engineering: a review[J]. Polymer Reviews, 2018, 58(1): 164-207（SCI源刊；IF: 5.375；署名单位：华中科技大学）

[2] **Suli Yang**, Jing Liu, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau. Principled Schedulability Analysis for Distributed Storage Systems using Thread Architecture Models. In: Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2018), Carlsbad, CA, USA, October 8-10, 2018. USENIX Association, 2018: 161-176（EI收录；署名单位：华中科技大学）

[3] **詹向红**, 李德新. 中医药防治阿尔茨海默病实验研究述要. 中华中医药学刊, 2004, 22(11): 2094-2096（署名单位：华中科技大学）

**专 利**

[1] 全部作者的姓名全称，本人的名字加粗. 专利题名. 专利国别，专利文献种类，专利号或申请号，专利申请日（专利权人：华中科技大学）

**专 利（举例）**

[1] 刘德林，李德奎. 多功能可擦写存储器. 中国，发明专利，202010575613.3，2020.3.20（专利权人：华中科技大学）

**标 准**

[1] 全部作者的姓名全称，本人的名字加粗. 标准题名. 哪种层次的标准，发表年

**科技奖励**

[1] 全部作者的姓名全称，本人的名字加粗. 题目. 国家级/省部级科技类奖，获奖年

[2] 全部作者的姓名全称，本人的名字加粗. 题目. 国际/国内竞赛类奖，获奖年

注：如果取得的各类研究成果均较多，可以将成果分别在不同附录中给出，这时附录的题目可以更改为“附录1 攻读博士学位期间发表的论文”，“附录2 攻读博士学位期间申请的专利”等。

# 附录2 攻读学位期间参加的科研项目

**1. 项目类型**

项目名称: 项目名称

项目编号: No. 88888888

起止时间: 2018年8月至2019年8月

担任角色：担任角色

# 附录3 中英文缩写对照表

3D Three-dimensional（三维）

CT Computer tomography（计算机断层层析成像）

MRI Magnetic resonance imaging（磁共振成像）

PET Positron emission computed tomography（正电子发射断层成像）

…

按字母顺序排列。

# 附录4 其它附录

可包括详细的公式推导、实验数据、计算程序、援引他人的原始资料、数据及其设备条件等。

如果本附录有内容，可以更改附录标题为相应的合适题目。

如果本附录没有内容，则删除本附录。