

# 基于群智能优化理论的若干 聚类改进方法及应用研究

周瑞红 著

 科学出版社

# 基于群智能优化理论的若干聚类 改进方法及应用研究

周瑞红 著

科学出版社

北京

## 内 容 简 介

伴随着大数据时代的到来,信息爆炸产生了海量数据,云计算、物联网以及移动互联网渗透到人们日常生活的各个领域,互联网金融、电子商务、信息科技等先进科技领域无时无刻不在产生数据。数据结构越来越复杂,数据形式越来越多样,深度挖掘海量数据中隐含的信息成为大数据时代经济管理发展的迫切需求。聚类是数据挖掘的重要方法,面对海量数据,提取有价值信息。本书是作者攻读博士期间科研成果的总结,全书共分为6章,重点针对密度峰值聚类方法进行若干理论改进与应用研究,并将其应用于实际生活各个领域,取得了令人满意的结果。

本书可供从事大数据研究的科研人员参考,也可作为高等院校相关专业高年级本科生和研究生的教材。

---

### 图书在版编目(CIP)数据

基于群智能优化理论的若干聚类改进方法及应用研究/周瑞红著.  
—北京:科学出版社,2018.5  
ISBN 978-7-03-056570-9

I. ①基… II. ①周… III. ①智能控制-最优化算法-研究  
IV. ①TP273

中国版本图书馆CIP数据核字(2018)第028928号

---

责任编辑:王 哲 / 责任校对:郭瑞芝  
责任印制:师艳茹 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

保定市 中画美凯印刷有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2018年5月第 一 版 开本:720×1000 1/16

2018年5月第一次印刷 印张:8 插页:2

字数:181 000

定价:59.00 元

(如有印装质量问题,我社负责调换)

# 前 言

伴随着大数据时代的到来，信息爆炸产生了海量数据，云计算、物联网以及移动互联网渗透到人们日常生活的各个领域，互联网金融、电子商务、信息科技等先进科技领域无时无刻不在产生数据。数据结构越来越复杂，数据形式越来越多样，深度挖掘海量数据中隐含的信息成为大数据时代经济管理发展的迫切需求。

聚类是数据挖掘的重要方法，可以作为发现数据集分布的独立工具，也可以用于对数据进行深入分析和其他方法的数据预处理过程。但几乎所有知名的聚类方法都需要手动输入关键参数，参数的选择完全依赖研究人员的先验知识，且这类参数对于聚类结果的影响是非常显著的。面对数据量大、结构复杂的数据集，人们根本无法做到完全准确地设定全局参数。

在统计学中，人们最常用的是 SPSS 统计分析软件中的二阶段聚类、快速聚类和层次聚类三种方法，并利用这三种聚类方法解决所有数据类型的聚类分析问题。面对多维、海量的复杂数据，这种方式只能得到粗略的计算结果，但对于管理者来说，准确的数据分析结果是正确决策的根本保证。对于不同的学习策略，需要使用不同的聚类方法。目前没有一种聚类方法能够准确地对所有类型数据进行分析。本书通过多种改进和创新密度峰值聚类方法，进一步丰富聚类理论，提供更多有效的聚类方法，满足研究人员分析复杂多维数据时的需要。密度峰值聚类方法是 2014 年在 *Science* 上提出的方法，具有聚类速度快、实现简单、可扩展性强等优点，适用于大数据环境下的数据分析。因此本书以密度峰值聚类方法为出发点，根据密度峰值聚类方法中截断距离参数  $d_c$  需要人为设定的不足，利用不同策略的群智能优化方法计算该参数，解决了密度峰值聚类方法中的参数估计问题。

本书的核心思想就是利用目前群智能方法中最新提出的果蝇优化方法和布谷鸟优化方法的特点，以密度峰值聚类方法的参数确定为主要研究对象，对果蝇优化方法和布谷鸟优化方法进行改进，提高原方法的科学性和鲁棒性，并利用改进的三种智能优化方法分别确定密度峰值聚类中的截断距离参数，提出三种改进的密度峰值聚类方法。

本书的主要内容和创新如下：

(1) 提出基于知识学习的改进果蝇优化方法，并将该方法应用于优化密度峰值聚类方法中的截断距离，进一步提出基于改进果蝇优化的密度峰值聚类方法。果蝇优化方法具有实现简单、参数少、效率高、应用能力强等特点，但该方法的完全随机

策略虽然简单，搜索过程却有较大盲目性，易陷入局部极值。针对此问题，本书引入知识学习策略，提高方法全局搜索能力和收敛速度。当果蝇优化方法陷入局部极值时，果蝇通过学习种群知识获得不同变异尺度。通过模拟仿真实验，利用多个测试函数验证了改进后的果蝇优化方法寻优能力更强。该方法优化密度峰值聚类方法的参数时，针对某些数据集比原聚类方法表现出更强的聚类能力。

(2) 提出基于动态发现概率的改进布谷鸟优化方法，并将该方法应用于优化密度峰值聚类方法中的截断距离，进一步提出基于改进布谷鸟优化的密度峰值聚类方法。布谷鸟优化方法具有参数少、鲁棒性好、全局搜索能力强等优势，但也存在方法寻优速度慢等缺陷。本书改进布谷鸟方法中的发现概率，使其与当前函数值变化关联起来，动态更新鸟巢位置，通过这样的改进完善原方法的完全随机特性。改进后的布谷鸟优化方法可以根据当前最优、最差鸟巢位置距离，有方向地控制发现概率的大小，使改进后的方法具有更快的寻优速度和更准确的收敛精度。通过模拟仿真实验，利用测试函数验证了改进后的布谷鸟优化方法寻优能力更强。该方法优化密度峰值聚类方法的参数时，针对某些数据集也表现出比原聚类方法更强的聚类能力。

(3) 提出智能选择优化方法。果蝇优化方法具有较强的局部搜索能力，同时布谷鸟优化方法的全局搜索能力更强。结合两者的优点，引入协同重组算子和基于混沌理论的变异因子。首先将初始化种群等分成双种群，并分别利用提出的改进果蝇优化方法和改进布谷鸟优化方法实行双种群协同进化策略，在每次完成迭代时利用协同重组算子让两个种群引进对方的全局最优解，实现种群间的实时信息交流，且对两种方法得到的最优解通过轮盘赌方式选取最优位置，并作为下一次迭代过程中布谷鸟种群的鸟巢位置和果蝇种群的全局最优食物源，达到协同进化，使方法具有局部串行挖掘和全局并行勘探的能力，全面提高方法的寻优精度、收敛速度和全局寻优性能。进而提出基于智能选择优化理论的密度峰值方法，且比原聚类方法具有更好的寻优精度和较强的聚类能力。

本书是作者在攻读博士期间的科研成果总结，是在国家自然科学基金项目(项目编号：61202306、1472049)的支持下完成的，受到吉林财经大学的资助。值此专著完成之际，作为吉林财经大学的一名教师，诚挚地向许正良教授的多年指导和吉林财经大学的鼎力支持表示感谢。参与本书撰写的还有刘翘铭、周丽媛等，在此一并表示感谢。虽然作者在本书的撰写过程中力求叙述准确、完善，但由于水平有限，书中难免存在欠妥之处，敬请广大读者批评指正。

# 目 录

## 前言

第 1 章 绪论	1
1.1 研究背景及研究意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 研究思路及研究内容	3
1.2.1 研究思路	3
1.2.2 研究内容	4
1.3 研究技术路线及研究方法	5
1.3.1 研究技术路线	5
1.3.2 研究方法	6
第 2 章 基础理论及国内外研究现状	8
2.1 群智能优化理论概述	8
2.1.1 群智能优化理论内涵	8
2.1.2 主要群智能方法	9
2.1.3 测试函数	13
2.2 聚类理论概述	20
2.2.1 聚类理论内涵	20
2.2.2 主要聚类方法	20
2.2.3 聚类评价指标	29
2.2.4 主要相似度量方法	31
2.3 国内外研究现状	33
2.3.1 研究现状	33
2.3.2 目前研究的不足	37
2.4 本章小结	37
第 3 章 基于改进果蝇优化的密度峰值聚类方法及应用	38
3.1 原始果蝇优化方法	38
3.1.1 理论模型	38

3.1.2	方法流程	38
3.2	基于知识学习的果蝇优化方法	40
3.2.1	理论模型	40
3.2.2	知识学习策略	41
3.2.3	方法流程	42
3.2.4	仿真模拟实验	43
3.3	基于改进果蝇优化的密度峰值聚类方法	49
3.3.1	理论模型	49
3.3.2	仿真模拟实验	49
3.4	基于改进果蝇优化的密度峰值聚类方法的实际应用	54
3.4.1	研究背景	54
3.4.2	数据分析	54
3.5	本章小结	57
<b>第 4 章</b>	<b>基于改进布谷鸟优化的密度峰值聚类方法及应用</b>	<b>58</b>
4.1	原始布谷鸟优化方法	58
4.1.1	理论模型	58
4.1.2	方法流程	60
4.2	基于动态发现概率的布谷鸟优化方法	62
4.2.1	理论模型	62
4.2.2	方法流程	62
4.2.3	仿真模拟实验	64
4.3	基于改进布谷鸟优化的密度峰值聚类方法	71
4.3.1	理论模型	71
4.3.2	仿真模拟实验	72
4.4	基于改进布谷鸟优化的密度峰值聚类方法的实际应用	76
4.4.1	研究背景	76
4.4.2	数据分析	76
4.5	本章小结	78
<b>第 5 章</b>	<b>基于智能选择优化的密度峰值聚类方法及应用</b>	<b>79</b>
5.1	基于智能选择的优化方法	79
5.1.1	理论模型	79
5.1.2	方法流程	80
5.1.3	仿真模拟实验	82

---

5.2	基于智能选择优化的密度峰值聚类方法	87
5.2.1	方法描述	87
5.2.2	方法流程	87
5.2.3	仿真模拟实验	89
5.3	基于智能选择优化的密度峰值聚类方法的实际应用	102
5.3.1	研究背景	102
5.3.2	数据分析	103
5.4	本章小结	104
第 6 章	结论与展望	105
6.1	研究结论及创新点	105
6.1.1	研究结论	105
6.1.2	研究创新点	106
6.2	研究局限和未来展望	107
6.2.1	研究局限	107
6.2.2	未来展望	107
	参考文献	109

彩图



# 第 1 章 绪 论

## 1.1 研究背景及研究意义

### 1.1.1 研究背景

伴随着大数据 (big data) 时代的到来, 信息爆炸产生了海量数据, 云计算、物联网以及移动互联网遍布人们日常生活的各个领域。信息量急速爆炸, 每天都会产生万亿至千万亿字节的数据量, 并且已经从 TB ( $1024\text{GB}=1\text{TB}$ ) 级别跃升到 PB ( $1024\text{TB}=1\text{PB}$ )、EB ( $1024\text{PB}=1\text{EB}$ ) 乃至 ZB ( $1024\text{EB}=1\text{ZB}$ ) 的级别, 亟待人们分析处理。据中国信息产业发展研究院数据, 我国云计算产业链规模已经达到 1 万亿元。百度每天处理的数据量将近 100PB, 相当于 5000 个国家图书馆的信息量的总和; Google 每天处理搜索诉求的数据量可以达到千万亿字节; 淘宝网每天产生数据量达到 10T, 绝大部分是由消费者、商家产生的, 也有物流公司和其他各种各样的信息; 滴滴出行全平台 (涵盖出租车、专车、快车、顺风车、代驾、巴士、试驾、企业版) 整体完成日订单 1000 万, 成为仅次于淘宝网的第二大互联网交易平台。滴滴出行每天处理数据 70TB, 每天做超过 90 亿次的路径规划, 每秒钟有 1000 多次用车需求, 这个数据量尤其是连续定位数据已经远超世界上任何一个地图公司。

随着社会的快速发展, 互联网金融、电子商务、信息科技等先进技术领域无时无刻不在产生数据。数据结构越来越复杂, 数据形式越来越多样, 深度挖掘海量数据中隐含的实际信息成为大数据时代社会发展的迫切需求。数据挖掘 (data mining) 作为数据库知识发现的一个重要手段, 是从海量的、不完全的、无规则的、有噪声的、模糊的数据中提取隐含在其中, 但又是潜在有用的信息和知识的过程。数据挖掘是一门交叉性学科, 继承了很多学科中成熟的工具和技术, 包括了人工智能、机器学习、神经网络、统计学等, 能够帮助各行各业的决策支持信息提供数据支撑和理论依据。

数据挖掘的方法主要指的是分类、估计、预测、相关性分组与聚类等。在数据挖掘的诸多技术中, 聚类 (clustering) 是最基本, 也是非常重要的数据挖掘方法之一, 认识数据、理解数据是聚类分析的基本功能, 它是无监督分类过程, 是人

工智能领域的研究重点。聚类分析起源于分类学，但是聚类不等同于分类。聚类与分类的不同在于，聚类所要求划分的类是未知的。聚类方法就是以相似性为基础，通过把物理或抽象对象的集合分成由类似的对象组成的若干个簇，通过聚类分析发现相似数据间的知识特征。因此由聚类所生成的簇是一组数据对象的集合，这些对象与同一个簇中的对象彼此相似，与其他簇中的对象相异。同时，聚类分析也应用到了工程领域如图像分割、文本分析等，也应用到了数据挖掘、客户分类、信息检索、基因识别、模式识别、Web 文档分类等领域。为了提高大数据统计及分析的效率，可以通过聚类方法对数据集进行处理，以此来减少数据集的维度，同时可以去掉相似的数据冗余。

在统计学中，人们最常用的是层次聚类、快速聚类和二阶段聚类，并利用这三种方法来解决一切维度数据的多元统计问题。面对多维、海量的复杂数据，这种计算方式无法保证对所有数据都能得到十分精确的聚类结果，但作为管理者来说，准确的数据分析结果是正确决策的根本保证。因此，开发出适应不同类型数据的聚类方法是顺应数据时代发展的必然要求。

### 1.1.2 研究意义

聚类分析方法一直是统计分析和机器学习领域的重要分支，但面对日益增长的复杂多维数据，仍然需要对传统聚类方法进行不断优化和改进，包括如何能够更有效地分析大规模数据，找出数据背后的潜在信息，以满足社会发展的迫切需求。在商业分析中，聚类分析可以帮助市场分析人员从消费者数据库中把不同的消费群体区分开来，并且概括出每一类消费者的消费模式或消费习惯。它作为数据挖掘中的一个模块，可以作为一个独立的数据分析工具来发现数据库中分布的一些深层次信息，并且概括出每一类的特点，或者把注意力放在某一个特定的类上进行深入的分析，而且聚类分析也可以作为数据挖掘方法中其他分析方法的一个预处理过程。

传统的聚类分析方法有系统聚类法、有序样品聚类法、动态聚类法、模糊聚类法、图论聚类法、聚类预报法等，这些传统的聚类方法能够比较成功地解决低维数据的聚类问题。但由于实际应用中数据的复杂性，现有的方法在处理多维数据和复杂数据时经常失效。许多在低维数据空间表现良好的聚类方法运用在多维空间上往往无法得到理想的聚类效果。

多维数据聚类分析是大数据时代聚类分析的一个重要研究内容，也是聚类技术的难点。技术的进步使得数据收集变得越来越容易，导致数据库规模越来越大、复杂性越来越高，如各种类型的商贸交易数据、Web 文档、基因表达数据等，它们的维度(属性)通常可以达到成百上千维，甚至更高。目前多维数据聚类分析在市场分析、信息安全、金融、娱乐、反恐等方面都有很广泛的应用。多维数据集

中存在大量无关的属性使得在所有维中存在簇的可能性几乎为零，且多维空间中的数据比低维空间中的数据分布更加分散，数据间距离几乎相等是普遍现象，而传统聚类方法是基于距离进行聚类的，因此在多维空间中无法基于距离来构建簇。密度峰值聚类(density peaks clustering, DPC)方法是 Rodriguez 和 Laio 于 2014 年在 *Science* 上发表的一种最新的聚类方法，具有方法简单、参数少、无序迭代、扩展性强、聚类效果好等特点，得到了广泛应用，但该方法的截断距离参数  $d_c$  定义模糊，仍有缺陷。

最优化问题利用数学符号进行表示之后，将其归纳为函数优化问题和组合优化问题两类。其中，函数优化问题是对一定区间内的连续变量进行处理和分析，而组合优化问题的对象则是对空间中离散的数据进行处理和分析。目前，一些精确方法和智能优化方法在解决具体优化问题方面发挥着重要作用。精确方法包括线性规划、整体规划、动态规划和分支定界等，其方法设计复杂，只适用于求解小规模数据问题。智能优化方法，包括遗传方法、蚁群方法、蜂群方法等，都是从任一解出发，通过对其邻域不断搜索和当前解的替换来完成具体的优化过程，并且根据搜索行为不同，又划分为局部搜索和全局搜索。根据生物进化行为，学者们提出了诸多群智能优化方法，通过学习、模拟和提炼自然界社会性群体行为，比如果蝇、蚁群、蜂群、布谷鸟等自然生物的群体活动提炼出具体的寻优方法，模拟和揭示其中的迭代过程，克服传统数学方法的局限，计算简单，具有明显的鲁棒性、自组织性、分布性和协作性。这些群智能优化方法的并行性特征和潜在分布式优势，为处理全局模型未知情形下的复杂分布式问题打下了坚实的基础。

本书为了解决密度峰值聚类方法的截断距离参数  $d_c$  需要人为设定的不足，首先对果蝇群智能优化方法、布谷鸟群智能优化方法进行改进，提出了基于知识学习的改进果蝇优化方法、基于动态发现概率的布谷鸟优化方法，以及引入协同重组算子和混沌变异算子的智能选择优化方法，并分别将这三种群智能方法应用到密度峰值聚类方法中的参数优化问题，一方面扩展了群智能方法的应用范围，另一方面也丰富了聚类方法的解决方案，并利用仿真模拟实验和现实数据验证三种聚类方法的有效性和可靠性。

## 1.2 研究思路及研究内容

### 1.2.1 研究思路

不同的学习策略，需要不同的聚类方法。因为没有一种聚类方法可以准确地对所有类型、所有规模数据进行分析。除了现有的传统群智能方法，近年

来, 国外学者还提出了多种新颖的群智能方法, 例如, 本书主要研究的果蝇优化方法、布谷鸟优化方法, 都是在生物进化的基础上逐步推演而来, 有很强的数学基础。

目前所提的聚类方法主要目的就是获取未知标签数据结构, 给出一种获取数据中价值信息的途径, 该技术需要的已知信息量较少, 正是基于这样的出发点, 聚类技术成为数据挖掘领域的重要方法。各个聚类方法的主要意图是一致的, 但是实现方法各有不同。密度峰值聚类方法正是利用类簇分布情形, 即类簇中心被具有较低局部密度的数据点包围, 且与具有更高密度的任何数据点有相对较大的距离, 以点密度及其相对距离属性来构建一个聚类模型, 这种聚类方法通过人为设定截断距离, 即邻域半径  $d_c$  的值, 来衡量该数据点相对于其他数据点的紧密程度。其理论基础简单, 但是聚类效果非常好。

优化方法是人们在工业生产、工程应用和运筹管理领域频繁处理的问题的重要解决办法。优化方法本质上就是一个搜索过程, 通过某种方式或某种规则找到问题的最佳解决方案。传统的优化方法计算复杂, 需要明确的目标函数信息, 只适合于解决小规模问题。当实际问题复杂程度高、建模复杂、约束条件多、维度高且拥有多个极值的情况下, 就需要专门的处理复杂优化问题的智能方法。本书的主要内容就是利用群智能优化方法寻优的特性, 比如果蝇优化方法 (fruit fly optimization algorithm, FOA)、布谷鸟优化方法 (cuckoo search optimization algorithm, CS) 等, 科学解决聚类方法中的参数设定问题。因为聚类方法中的搜索半径都是根据研究人员的先验知识确定, 在大数据环境下, 这种设定参数的方法过于主观, 严重影响了聚类结果的科学性、准确性和鲁棒性。

根据基于计算智能的基本理论, 本书针对密度峰值聚类方法中存在的不足之处, 结合群智能方法的特点, 尤其是果蝇群智能优化方法和布谷鸟群智能优化方法的特点, 以密度峰值聚类方法为改进对象, 对果蝇群智能优化方法和布谷鸟群智能优化方法进行改进, 提高原方法的科学性和鲁棒性。通过若干数据集的实验验证, 基于改进群智能优化方法的密度峰值聚类方法收敛速度有了更好的改善, 适应性也得到了大大增强, 智能化分析的能力大幅度提高, 应用前景广泛, 比过去传统的统计聚类方法能更快、更准确地揭示海量数据间的复杂关系, 提高了数据聚类计算的实用性, 有效地提升了数据挖掘的效率和质量。

### 1.2.2 研究内容

本书具体内容安排如下:

第 1 章为绪论, 首先介绍了本书研究的背景及研究意义。对主要研究内容进行梳理, 对结构安排加以介绍, 并简要阐述了本书的三个主要创新之处。

第2章介绍群智能优化理论,对经典的群智能优化方法,如遗传方法、蚁群方法、粒子群方法的方法流程进行说明,对果蝇优化方法、布谷鸟优化方法加以详细介绍。对本书主要改进的果蝇优化方法和布谷鸟优化方法的国内外发展现状进行梳理。

第3章介绍原始果蝇优化方法的理论模型,并提出基于知识学习的改进果蝇优化方法,引入知识学习策略,提高原始果蝇优化方法全局搜索能力和收敛速度。利用测试函数验证了改进后的果蝇优化方法寻优能力更强。然后将该方法应用于优化密度峰值聚类方法中的截断距离 $d_c$ ,进一步提出基于改进果蝇优化的密度峰值聚类方法。针对某些数据集该方法也表现出更强的聚类能力。最后把改进后的聚类方法应用于现实数据,证明该方法的实用性。

第4章介绍原始布谷鸟优化方法的理论模型,提出基于动态发现概率的改进布谷鸟优化方法,并将该方法应用于优化密度峰值聚类方法中的截断距离,进一步提出基于布谷鸟优化理论的密度峰值聚类方法。通过模拟仿真实验,利用测试函数验证了改进后的布谷鸟优化方法寻优能力更强。该方法优化密度峰值聚类方法时,针对某些数据集也表现出更强的聚类能力,并结合现实数据进行应用。

第5章介绍智能选择优化方法的基本理论。结合果蝇优化方法具有较强的局部搜索能力和布谷鸟优化方法较强的全局搜索能力,引入协同重组算子和基于混沌理论的变异因子,让两个种群引进对方的全局最优解,对两种方法得到的最优解通过轮盘赌方式选取最优位置,并作为下一次迭代过程中布谷鸟种群的鸟巢位置和果蝇种群的全局最优食物源,进而达到协同进化,利用若干不同的测试函数和数据集检验方法的质量。得到的最优解作为密度峰值聚类方法中的截断距离,产生新的改进聚类方法。

第6章总结本书提出的三种改进方法及存在的主要问题,根据之前的分析得出结论,并展望未来的研究重点和方向。

## 1.3 研究技术路线及研究方法

### 1.3.1 研究技术路线

本书的技术路线如图1.1所示。

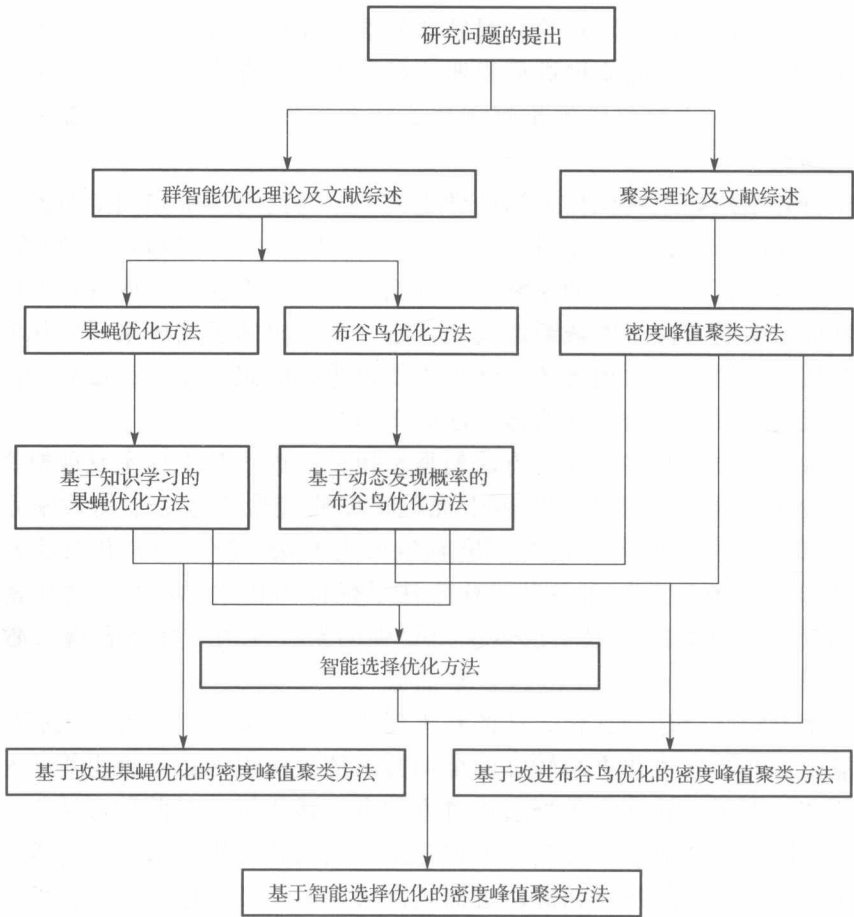


图 1.1 技术路线图

### 1.3.2 研究方法

本书在研究过程中，始终注重通过二手数据分析，梳理目前的群智能优化方法，掌握关于果蝇优化方法和布谷鸟优化方法，以及密度峰值聚类方法的国内外研究现状。通过模拟仿真实验和测试函数检验书中提出的基于知识学习的果蝇优化方法、基于动态发现概率的布谷鸟优化方法、基于智能选择优化方法，并分别将三种改进方法优化密度峰值聚类中的截断距离参数  $d_c$ 。

(1) 综合研究与比较研究相结合。

综合研究注重理论和知识在多个层面、多个角度的综合集成，通过二手资料分析，了解本书主要研究的智能方法及聚类方法的国内外研究进展，分析现有研究的不足；在本书对每个改进方法的仿真模拟实验中，通过比较研究方式，比较

改进前后的群智能优化方法，比较改进后的密度峰值聚类方法与统计学中常用的快速聚类、二阶段聚类法和层次聚类的聚类效果，以此说明改进后的方法的实用性和科学性。

(2) 理论与模拟仿真实验相结合。

理论研究是对某一领域的知识进行探讨，形成基本的理论架构，对该领域具有方向性的指导意义。本书对群智能优化理论和聚类理论进行梳理，分析现有群优化理论和聚类理论的不足，提出主要改进方法。通过模拟仿真实验，证明改进后的优化方法比原方法有更强的先进性。

(3) 规范研究与实证研究相结合。

规范研究是指运用演绎和归纳方法，注重从逻辑方面概括“应该怎样”或“应该怎样解决”的方法。其主要特点是在进行分析以前，要先确定相应的准则，然后再依据这些准则来分析判断研究对象目前所处的状态是否符合这些准则，如不符合，则计算其偏离程度，研究如何调整。实证研究是指以事实结果为标准，验证与衡量理论或观点、假说的正确性，这种方法主要回答“是什么”的问题。

## 第 2 章 基础理论及国内外研究现状

### 2.1 群智能优化理论概述

#### 2.1.1 群智能优化理论内涵

在自然界中，动物经常以集体的行为方式外出觅食。而且在这些生物群体中，单个个体所表现出来的仅仅是简单且缺乏智能的行为，每个独立个体所做的行为都是相近的，但在某种进化机制的引导下，这些个体组成的群体却可以通过个体之间的交互关系发挥作用，个体以及可能的个体智能无法独立完成任务通过群体的共同努力可以完全实现。适者生存，这是自然进化的结果，对它们的实际生存和进化有着十分重要的影响。1999 年，群智能 (swarm intelligence) 的概念被正式提出，任何由社会昆虫群体和动物群体的集体行为启发而提出的方法和分布式问题的解决方案，都称为群智能研究，由其演化出来的方法称为群智能优化方法 (swarm intelligence optimization algorithm)。由众多非智能简单个体组成的整个群体，通过相互间的简单合作表现出整体智能行为的特性。目前，对群智能的研究尚处于初级阶段，但是它正不断受到国际智能计算研究领域学者们的关注，并逐渐成为一个新的重要研究方向。它是众多非智能的简单个体组成的群体之间通过简单合作而表现出来的整体智能性。Benabeau 等于 2000 年提出对昆虫行为进行研究，为那些在分布式控制和优化方法领域的计算机科学家们提供了非常有价值的参考信息，而且这些方法被充分应用到经济、工程设计、交通管理、生产调度等领域中。

群智能优化理论以生物社会性为依托，模拟若干独立个体组成的群体以及个体之间的相互行为形成的生物社会性。如大雁在空中集体飞行时会自动排列成“一”字形或“人”字形；大量蝙蝠同时在黑暗的洞穴中快速飞行，不相互碰撞。对于这些生物之间和谐相处现象的一种合理解释就是，群体中的所有独立个体都遵守某些特定的社会性规则，当它们按照这些社会规则相互发生作用时就会自觉地发生上述行为。这种模拟系统可以通过局部信息的收集和分析协调整个群体的行为，并依此来解决难以建立有效形式化模型的问题，也可以有效解决传统优化方法难以有效解决或者根本无法解决的问题。群智能优化理论逐渐成为人工智能领域的又一研究热点。

目前研究群智能的方法多是以多主体系统为出发点来进行的。该观点假定多



主体系统中的每个个体都具备感知周围环境的能力,包括自身和其他主体对环境的改变,各主体间通过环境变化来彼此间接通信。国内外学者也从多主体系统的角度分析群智能的性能特点,可以相互通信交流是群智能的共同特征,每个主体都能智能地存取所需信息,是一种分布式的计算模式。在原有群智能的研究与发展的基础上,学术界又提出了多种群智能优化方法,最典型的有遗传方法、蚁群方法、粒子群方法以及蜂群方法等,为解决优化问题提供了新思路。

人们通过对自然界中的一些如蚂蚁、蜜蜂等昆虫的观察,提出了群智能的概念。单只昆虫的智能性表现并不明显,几只昆虫凑到一起就能一起往巢穴搬运路上遇到的大型食物。如果是一群昆虫,它们就能协同工作,建立起坚固、漂亮的居所,一起抵御危险繁育后代,这种群居性生物的群智能充分体现出一种群智能能力。1994年 Millonas 提出生物的群智能行为遵循5条基本原则:群体能够完成简单的空间和时间计算,能够响应环境中的品质因子,行动范围宽泛,每次环境变化时不会改变自身的行为,群体主动在适当的时候改变自身的行为。这些原则充分说明要想实现群智能,智能主体必须具备自主性、反应性、学习性和自适应性等智能特性。但这并不代表群体中的每个个体都相当复杂,事实恰恰与此相反。就像单只昆虫智能能力不高一样,组成群体的每个个体都只具有简单的智能能力,它们通过相互之间的协同合作表现出复杂的整体智能行为。因此,群智能的核心是指由众多简单个体组成的群体能够通过相互之间的简单合作来实现某一功能或完成某一任务。其中,“简单个体”是指单个个体只具有简单的能力或智能,而“简单合作”是指个体和与其临近的个体进行某种简单的直接通信或通过改变环境间接与其他个体通信,从而完成相互影响、协同动作。鲁棒性是群智能的显著特点,不会因为某个或某几个个体的行为而影响整个群体对某些问题的解决能力。群体中的每个个体都具有通过个体之间间接通信改变环境的能力。随着个体数目的增加,群智能的实现越来越容易。群体具有自组织性,能够通过简单个体的交互过程表现出来整体的智能性(emergent intelligence)。

### 2.1.2 主要群智能方法

#### (1) 遗传方法

遗传方法(genetic algorithm, GA)是 Holland 教授于 1975 年提出来的进化方法。该方法模拟达尔文自然生物进化论中的自然选择和遗传学中的遗传变异理论模拟生物进化过程的计算模型,即遵循物竞天择、优胜劣汰的自然进化法则,通过模拟自然进化过程,反复使用遗传学的基本操作,不断生成新的种群,不断进化,同时全局并行搜索群体中的最优解。它的特点是对参数进行编码运算不需要有关体系的任何先验知识,沿多种路线进行平行搜索,不会落入局部较优的陷阱。