



邱春艳◎著

群体智能 算法改进及其应用



科学出版社

大数据分析与应用丛书

群体智能算法改进及其应用

邱春艳 著

科学出版社

北京

内 容 简 介

群体智能算法是模拟自然生物种群智能行为的优化方法,具有良好的寻优性能,因此群体智能算法在求解大规模复杂问题时具有较高的效率。本书在群体智能的基础上,针对智能优化算法、聚类算法、复杂网络和朴素贝叶斯分类进行理论研究和应用研究。本书重点研究群体智能算法中的人工蜂群算法理论改进和应用研究,提出基于人工蜂群的密度峰值聚类算法,并将其应用到民营上市公司的聚类分析中。另外,本书还提出基于复杂网络和朴素贝叶斯分类的人工蜂群算法,使用复杂网络理论,降低供应链网络结构的复杂度;引入朴素贝叶斯分类,大幅度地加快算法的寻优速度,并将其应用于供应链网络优化决策中,都取得了令人满意的结果。

本书可供计算机技术与工程、信息技术等领域,以及关注群体智能优化、供应链管理的工程技术人员和科研人员参考,也可供高等院校智能优化、数据挖掘等专业的师生阅读。

图书在版编目(CIP)数据

群体智能算法改进及其应用/邱春艳著. —北京:科学出版社, 2019.3
(大数据分析与应用丛书)

ISBN 978-7-03-060795-9

I. ①群… II. ①邱… III. ①人工智能-算法-研究 IV. ①TP18

中国版本图书馆 CIP 数据核字(2019)第 044277 号

责任编辑:王喜军 常友丽 / 责任校对:韩 杨

责任印制:吴兆东 / 封面设计:壹选文化

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>

北京九州迅驰传媒文化有限公司 印刷

科学出版社发行 各地新华书店经销

*

2019 年 3 月第 一 版 开本:720×1000 1/16

2019 年 3 月第一次印刷 印张:6 3/4

字数:160 000

定价:88.00 元

(如有印装质量问题,我社负责调换)

丛书编辑委员会

主任委员：王丽敏

副主任委员：刘 洋 邱春艳

委 员：（按姓名首字笔画排序）

丁 楠 于卓熙 马知行 王 帅 王 楠 刘 颖

孙铁铮 李 军 李睿智 沐光雨 张池军 相薷薷

姜建华 曹忠波 韩佳玲

秘 书：马 艳

丛 书 序

在这个信息爆炸的时代，数据正以前所未有的规模急剧增长，海量数据的收集、存储、处理、分析以及信息服务成为全球信息技术发展的主流。从1997年美国航空航天局研究员迈克尔·考克斯和大卫·埃尔斯沃斯首次使用“大数据”这一术语到现在，“大数据”已成为当下最火热的IT行业词汇。2008年9月 *Nature* 推出了 *Big Data* 专刊，“大数据”自此成为政府、学术界、实务界共同关注的焦点。IBM 公司提出了大数据的 5V 特点：大量（volume）、高速（velocity）、多样（variety）、低价值密度（value）和真实性（veracity）。大数据的特点驱动着大数据分析技术与方法应运而生。

大数据研究之所以备受关注，源于大数据隐含着巨大的社会、经济、科研价值，在物联网、舆情分析、电子商务、健康医疗、生物技术、金融信息、人工智能等各个领域具有广泛的应用。因此，大数据分析作为深度挖掘数据价值的关键手段，引起了各行各业的高度重视。有效地组织和使用大数据，将对社会经济和科学研究产生巨大的推动作用，同时也孕育着前所未有的机遇。目前，大数据已经成为与自然资源、人力资源一样重要的战略资源，是一个国家数字主权的体现。大数据时代，如何提高和改进人们从海量和复杂的数据中获取知识的能力，已经成为各国学者学术研究的焦点。

在此背景下“大数据分析与应用丛书”出版。这套丛书是对现有大数据相关著作的补充与完善，基本覆盖大数据分析、人工智能、数理统计、机器学习、电子商务等各个领域，融合了大数据学术领域的相关技术、算法与大数据应用领域的具体实例和方法，是一套大数据分析与应用相结合的佳作。本丛书可作为统计学、管理学、计算机科学、金融信息、大数据、电子商务、信息系统等专业的本科生、研究生进行数据挖掘、机器学习、人工智能、数理统计等相关研究时的参考。

“大数据分析与应用丛书”由吉林财经大学数据科学与大数据技术专业、计算机科学与技术专业、电子商务专业、信息管理与信息系统专业的专任教师和吉林省互联网金融重点实验室、吉林省商务大数据研究中心科研人员撰写，还有一批有影响力的学者组成了编委会负责书稿审校。相信丛书出版后将对本领域的教师、科研人员、工程师、管理人员、学生和爱好者有所裨益，为大数据分析与应用领域的研究与发展贡献一份力量。

本丛书得到国家自然科学基金项目（项目编号：61472049、61572225、61402193）、国家社会科学基金项目（项目编号：15BGL090）的资助，丛书编委会和全体作者在此表示衷心的感谢。

“大数据分析与应用丛书”编委会

2018年9月5日

前 言

供应链管理作为企业的左膀右臂，对企业能否健康发展起着重要作用，是给企业带来巨大收益的竞争利器，备受企业决策者的关注。然而，随着互联网金融的迅猛发展，越来越多的制造企业在采购原材料时不再受制于地域的限制。全世界满足要求的供应者蜂拥而来，供应商数量急剧增加，企业决策者如何正确、迅速地从中选取满足自身利益要求、稳定、高质量的供应商变得非常困难。

智能优化算法能够从大量的原始数据中挖掘出隐含的、有价值的信息，快速求解复杂的优化问题。供应链网络结构复杂的特征决定了应用智能优化算法在供应链网络的管理和决策中具有较强的可行性和现实性。

本书通过对现今供应链中各供应商数据和网络拓扑结构的分析，利用智能优化算法和聚类方法，针对不同问题提出改进的算法，并对一家推土机制造企业的供应链网络进行优化和决策，主要内容如下。

(1) 针对密度峰值聚类算法对截断距离参数依赖性过强、数据点之间聚类结果过度影响所导致的类簇间数据点聚类结果的不确定性和缺乏客观性的问题，提出基于人工蜂群的密度峰值聚类算法，并将其应用到民营上市公司的聚类分析中。

(2) 针对供应链网络结构复杂、数据量大、求解速度较慢等问题，提出

基于复杂网络和朴素贝叶斯分类的人工蜂群算法；使用复杂网络理论，降低供应链网络结构的复杂度；引入朴素贝叶斯分类，大幅度地加快算法的寻优速度，并将其应用于供应链网络优化决策中。

本书针对智能优化算法中的人工蜂群算法存在的问题，结合密度峰值聚类、复杂网络、朴素贝叶斯分类等算法对其进行改进。结果表明，本书提出的方法能够较为有效地对供应商进行筛选，为企业决策者寻求更有效的供应链优化方案。

本书在吉林省社会科学基金项目“云安全视角下互联网金融数据风险分析与控制研究”（项目号：2017BS28）和“面向能耗感知的商业云 workflow 智能优化研究”（项目号：2018B79）、吉林省教育科学“十三五”一般规划课题“地方院校高级应用型人才校企协同培养模式与实践研究”（项目号：GH170336）、“基于‘双主体、三平台、四驱动’校内外联动机制的高校创业教育生态系统构建研究”（项目号：GH180288）、吉林省教育厅“十三五”科学技术项目（项目号：JJKH20180465KJ）、长春市哲学社会科学规划项目“长春市民幸福感测度及影响因素研究”（项目号：CSKT2018ZX-010）、吉林财经大学博士基金项目“在线医疗健康服务中用户隐私信息披露行为与保护机制研究”（项目号：2018B15）等的资助和支持下完成。

由于作者水平有限，加之智能研究领域纵深宽广，书中难免有种种考虑不周之处，诚请广大读者批评指正。

邱春艳

2018年7月于长春

目 录

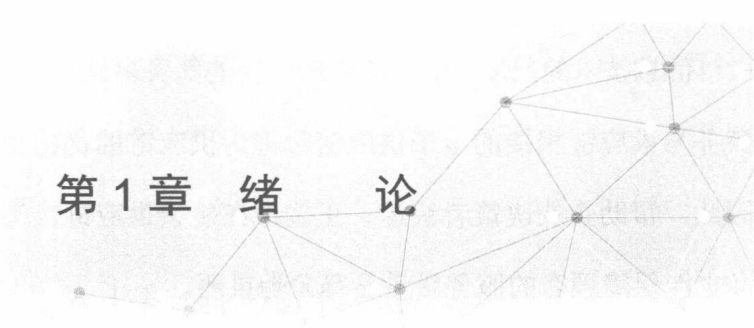
丛书序

前言

第 1 章 绪论	1
1.1 研究背景	1
1.2 研究目的	3
1.3 研究意义	6
1.4 本书章节安排	9
参考文献	10
第 2 章 国内外研究现状	12
2.1 群体智能优化算法发展现状	12
2.2 聚类技术在数据挖掘领域的研究现状	16
2.3 多目标优化方法及其在供应链优化领域的应用现状	18
2.4 本章小结	20
参考文献	21
第 3 章 群体智能相关理论研究	26
3.1 智能优化算法	26
3.2 聚类算法	28
3.3 复杂网络	29
3.4 朴素贝叶斯分类	30
3.5 本章小结	31

参考文献	31
第4章 基于人工蜂群优化的密度峰值聚类	34
4.1 引言	34
4.2 密度峰值聚类算法的改进	35
4.2.1 计算数据点的密度及生成决策图	37
4.2.2 执行初始聚类	38
4.2.3 识别类簇间数据点	39
4.2.4 初选类簇间数据点的类簇标号	40
4.2.5 判定类簇间数据点的类簇标号	41
4.2.6 完成聚类	44
4.3 实验结果	44
4.3.1 类簇间数据点的自动识别	44
4.3.2 任意形状数据集的类簇中心点和类簇数目的自动识别	48
4.3.3 不同形状和大小数据集的有效聚合	49
4.3.4 分类效果评价	51
4.4 结果分析	52
4.5 本章小结	54
参考文献	54
第5章 基于人工蜂群优化的密度峰值聚类算法的应用	56
5.1 引言	56
5.2 数据来源及预处理	57
5.3 基于人工蜂群的密度峰值聚类算法	59
5.4 聚类结果分析	60
5.5 本章小结	61
参考文献	62
第6章 基于复杂网路和朴素贝叶斯分类的人工蜂群算法	63
6.1 引言	63

6.2	模型建立	66
6.2.1	问题描述	66
6.2.2	解的表达	69
6.3	基于复杂网络的人工蜂群算法	71
6.4	基于复杂网络和朴素贝叶斯分类的人工蜂群算法的实现	73
6.5	推土机制造企业的供应链网络优化	74
6.5.1	全局最优解集的搜索能力	80
6.5.2	搜索全局最优解集的速度	89
6.6	算法结果分析	91
6.7	本章小结	93
	参考文献	93



第 1 章 绪 论

1.1 研究背景

随着全球经济的迅猛发展和经济一体化进程的加快，市场竞争日趋激烈，产品从研发到淘汰的生命周期越来越短，顾客对产品的要求越来越高，这些因素都迫使广大企业开始关注它们的供应链，希望通过提高供应链的管理水平来获得更多的利润。而供应链管理的出现，给企业间的合作与竞争带来了根本性的变化。供应链管理就是从全局的角度对供应链上的信息流、物质流以及资金流进行控制和调度，最终达到将消费者所需的产品在准确的时间并按照消费者所需的数量、所要求的质量以及正确的状态送到消费者所要求的地点，使得总成本最小。21 世纪的市场竞争重点由企业与企业之间的竞争转化为供应链与供应链之间的竞争，供应链的管理与优化已经成为企业核心竞争力的重要组成部分。随着企业规模的不断扩大，市场竞争态势不断变化，企业供应链的结构与运营模式总是处于动态调整之中，因此对供应链网络进行优化，提高整个供应链的运行效率，就显得很有必要。

信息技术的发展,带来了一种全新的商务模式——电子商务,而实现电子商务与供应链相接的电子供应链管理为供应链的优化提供了一种有效手段。因此,帮助企业决策者快速、正确地在复杂供应链决策中既能保障企业的利润也能保障顾客的服务质量显得尤为重要。

关于供应链比较常用的提法是“供应链是由物料获取并加工成中间件或成品,再将成品送到用户手中的一些企业和部门构成的网络”^[1-3]。供应链管理关注的重点已由原先的局部过程优化转向对整个供应链的统一协调,供应链的各节点企业、各环节已不再独立,而是以最大化地满足最终顾客需求为共同目标。

在这样的背景下,供应链优化与协调得到了众多学者的青睐以及企业界人士的广泛关注。近年来,行为因素对供应链优化与协调的影响成了研究的热点。人们对供应链的研究主要集中在供应链的设计、分析、协调、管理、优化和重组等方面。供应链管理的核心思想在于流程的优化和系统整体性能的提高,一些有效的方法逐渐被用来解决供应链优化问题,其中包括基于规则的系统、线性规划、约束传播和人工蜂群算法、粒子群算法等智能优化算法。大量研究实验表明,这些优化算法都能实现供应链决策优化问题的求解。然而,随着电子商务完全融入人们的日常生活、社会生产之中,供应链网络上的数据具有动态性、数据量大以及数据间关系复杂等特点,大大增加了供应链网络结构的复杂程度。因此,直接使用原始算法求解该问题的求解速度逐渐降低,很难达到预期效果。决策者更难在数据量如此庞大、数据间关系如此复杂的情况下快速、正确地找到既能保障企业利润又能保障服务质量

的供应链网络。因此，本书主要采用聚类及群体智能算法以供应链多目标优化为研究对象进行研究。

1.2 研究目的

本书主要采用聚类及群体智能算法对供应链多目标优化进行研究。首先，对相关理论进行讨论研究，探讨这些智能算法的优缺点，以及本书选择这些算法的原因；其次，分析基于人工蜂群的密度峰值聚类；最后，把基于人工蜂群的密度峰值聚类算法应用到从国泰安数据库中获取的 79 家民营上市公司的相关数据中，实现不同经营状况民营上市公司的分类。

(1) 探讨分析相关智能优化算法及本书采用这些算法的原因。

群体智能优化算法 (swarm intelligent optimization algorithms)^[4]的思想来自于仿生学，通过模拟大自然中动物的群体性行为，使个体根据群体的协调合作达到最优的状态。例如，鱼类的游动过程，蚂蚁的觅食过程，鸟群的飞行过程等。群体智能优化算法在搜索过程中，每个个体既能以群体的方向作为指导进行搜索，又有个体本身的独立自主性，跟传统的数值优化算法相比较，可以避免陷入局部最小值，更加易于找到全局最优解，也更加简便易行。因此，群体智能优化算法被国内外众多研究者不断地进行研究和探索，粒子群、蚁群以及人工鱼群技术是提出时间相对较早并且应用较为广泛的群体智能算法。人工蜂群算法是继这些之后的又一种新的优化技术，本书把人工蜂

群算法应用到供应链多目标优化中并进行研究。

聚类已经被深入地研究了许多年,聚类算法分为基于划分的聚类算法、基于层次的聚类算法、基于密度的聚类算法和智能型聚类算法 4 类。基于密度的聚类算法根据数据的密度进行聚类^[5],将高密度区域划分为一个类簇,能够识别任意形状类簇,典型代表是具有噪声的基于密度的聚类(density-based spatial clustering of applications with noise, DBSCAN)算法^[6]。但是 DBSCAN 算法需要确定邻域密度阈值和半径多个参数,聚类结果受参数影响比较大,文献[7]提出一种快速搜索发现密度峰值的聚类算法,能够快速、简洁地确定类簇。因此本书使用的是基于密度的算法中的基于密度峰值的聚类(density peaks clustering, DPC)算法。

复杂网络(complex network)理论可以用来描述大多数含有相互作用的实际系统,比如股票指标间的关联性。大量包含多个体和多个体相互作用的系统都可以抽象为复杂网络,其中每一个个体对应网络的一个顶点,个体之间的联系或相互作用对应连接顶点的边。供应链网络具有复杂网络的特征,Edward 等^[8]对已有的关于供应链网络属性的观点进行综合整理,将现实供应链网络属性映射到复杂网络模型中,从理论和现实相结合的角度深入分析、阐述了供应链网络的复杂网络属性,认为现实供应链网络类似无标度网络,具有短特征路径长度、高集聚系数、幂律连接分布特征。国内学者叶笛^[9]讨论了将复杂网络理论引入供应链网络研究领域的重要性,剖析了供应链的复杂网络特征,认为可利用复杂网络理论深入分析供应链系统整体运行规律和宏观行为、供应链网络的特征以及供应链网络动态生长演化规律。

分类是数据挖掘和机器学习中的一个重要研究领域^[10]，主要依靠对已经确定类别的数据进行分析，建立分类器，然后通过分类器对未知数据的类别进行判断^[11]。目前，已经存在大量发展成熟的分类算法，比较经典的算法有近邻分类算法^[12]、贝叶斯分类算法^[13]、支持向量机分类算法^[14]、神经网络分类算法^[15]、决策树分类算法^[16]和关联规则分类算法^[17]。朴素贝叶斯分类算法是最为经典的分类算法。朴素贝叶斯法分类算法是基于贝叶斯定理与特征条件独立假设的分类算法，所需估计的参数很少，对缺失数据不太敏感，算法也比较简单。将朴素贝叶斯分类中的贝叶斯定理与人工蜂群算法进行结合，可以加快人工蜂群寻优速度。

(2) 提出基于人工蜂群优化的密度峰值聚类 (bee-based density peaks clustering, BeeDPC) 算法。

聚类分析所使用的数据集中，各个类的密集程度往往不尽相同，甚至差别很大。大多数现有的聚类算法都是致力于如何发现任意形状和大小的类，但很难有效地处理密度差别较大的数据集。多密度聚类是数据挖掘中的一项重要技术，其目标是发现多密度数据集中的不同密度的数据对象，将其分离出来进行数据的处理和分析。因此，多密度聚类算法的研究为将复杂的海量数据转换为需要的信息提供了强有力的分析手段，它们今后将会在更多的领域中发挥作用。因此，本书提出 BeeDPC 算法，以聚类原则为基本思想，改进 DPC 算法，能够更好地实现类簇间数据点的自动识别和合理聚类。

(3) BeeDPC 应用到具体企业数据中，验证算法良好的实用性。

经过 20 多年的改革和发展，民营经济克服了基础薄弱和先天不足等劣势，

已成为国民经济的重要组成部分和国民经济中最为活跃的经济增长点。为了验证上面提出的 BeeDPC 算法的实用性,本书从国泰安数据库中获取 2015 年 12 月 31 日到 2017 年 12 月 31 日部分民营上市公司的相关数据,并使用基于人工蜂群的密度峰值聚类算法对该数据聚类分析。通过聚类分析,为供应链网络中心企业的决策者在选取备选供应商时提供可靠的依据,算法具有广泛的适用性。

1.3 研究意义

本书对供应链多目标优化展开研究,具有的理论意义和实践意义如下。

1. 理论意义

第一,提出基于人工蜂群的密度峰值聚类算法,降低了原始算法对截断距离 d_c 值的依赖性和数据点间的连带效应,使类簇间数据点的分类更加合理。将人工蜂群算法引入密度峰值聚类算法中,改变密度峰值算法原有的聚类原则,引入三角形稳定思想,充分挖掘数据点的潜在信息,通过比较待分类数据点、最高密度数据点、较高密度数据点三者距离上的相互关系,得到类簇间数据点可能的聚类结果集合,使用 $CDbw$ 值为目标函数的蜂群算法求得类簇间数据点聚类的最优解,得到全部数据点的聚类结果。仿真实验结果表明,基于人工蜂群的密度峰值聚类算法在保留原始算法的优良性能上能够