大数据特征降维

——粗糙集特征选择的群智能方法及应用研究

胡玉荣 著

中国水利水电出版社
www.waterpub.com.cn

大数据特征降维

——粗糙集特征选择的群智能方法及应用研究

胡玉荣　著

中国水利水电出版社
www.waterpub.com.cn
·北京·

内 容 提 要

本书从高维大数据的特征降维出发，指出大数据时代粗糙集特征选择面临的挑战，介绍了群智能算法的独特优势和存在的问题，对粗糙集和群智能的理论与经典算法进行了总结归纳并提出一种基于群智能和粗糙集的特征选择框架，依据此框架设计相关特征选择算法，应用于银行个人信用评分系统与高维数据集进行特征降维。

本书可供从事机器学习和大数据挖掘的高校教师、研究生、科研院所的科研人员及有关工程技术人员使用。

图书在版编目（CIP）数据

大数据特征降维 ：粗糙集特征选择的群智能方法及应用研究 / 胡玉荣著. -- 北京 ：中国水利水电出版社，2019.1

ISBN 978-7-5170-7363-5

Ⅰ. ①大… Ⅱ. ①胡… Ⅲ. ①数据采集－研究 Ⅳ. ①TP274

中国版本图书馆CIP数据核字(2019)第016070号

策划编辑：杨庆川　责任编辑：杨元泓　加工编辑：王开云　封面设计：李　佳

书　名	大数据特征降维——粗糙集特征选择的群智能方法及应用研究 DASHUJU TEZHENG JIANGWEI——CUCAOJI TEZHENG XUANZE DE QUNZHINENG FANGFA JI YINGYONG YANJIU
作　者	胡玉荣　著
出版发行	中国水利水电出版社 （北京市海淀区玉渊潭南路 1 号 D 座　100038） 网址：www.waterpub.com.cn E-mail：mchannel@263.net（万水） sales@waterpub.com.cn 电话：（010）68367658（营销中心）、82562819（万水）
经　售	全国各地新华书店和相关出版物销售网点
排　版	北京万水电子信息有限公司
印　刷	三河市元兴印务有限公司
规　格	170mm×240mm　16 开本　12.25 印张　230 千字
版　次	2019 年 4 月第 1 版　2019 年 4 月第 1 次印刷
定　价	55.00 元

前　言

伴随着科技新浪潮，计算机和互联网技术日益普及，大数据时代已悄然来临。大数据正在成为重要的战略资源，对大数据进行分析与挖掘是其发展的关键。如何降低数据的维度、避免“维数灾难”是数据挖掘工作的重中之重。随着描述数据的特征维数越来越高，大量针对降维提出的粗糙集特征选择算法面临严峻挑战。群智能方法是一种新型智能优化方法，具有协作性、简单性和分布性等特点，已在粗糙集特征选择中崭露头角，并彰显出独特优势。然而，群智能方法还有一些主要问题需要解决：早熟问题广泛存在不容忽视；对于大规模优化问题，算法后期容易出现停滞；参数多凭经验设置，对具体问题和应用环境依赖性大。

因此，如何对高维大数据进行特征选择是一项充满挑战的艰巨任务。本书针对群智能的这些问题及其在粗糙集特征选择中的应用进行了研究。本书第 1 章为绪论，介绍特征选择的概况、研究背景和研究现状。第 2 章针对粗糙集和群智能的理论和经典算法进行了总结归纳，并提出一种基于群智能和粗糙集的特征选择框架。第 3 至 5 章，依据此框架，提出三种基于群智能和粗糙集的特征选择算法。第 6 至 7 章，将三种算法应用于银行个人信用评分系统与高维数据集进行特征降维。第 8 章进行总结和展望。

本书理论与应用相结合，力求成为从事机器学习和大数据挖掘的高校教师、研究生、科研院所的科研人员及有关工程技术人员的参考书。全书由作者独撰，共二十三万字。本书的编写受到了荆楚理工学院引进人才科研启动金项目“面向高维大数据特征降维的群智能优化算法及相关问题研究”（编号：QDB201605）的资助。在此，表示感谢！

由于作者水平有限，书中难免存在不足之处，望广大读者给予批评和指正。

作者

2018 年 8 月

目　　录

第 1 章　绪论

1.1　本书研究背景

伴随着科技新浪潮，计算机和互联网技术日益普及，大数据时代已悄然来临。大数据即海量信息，由于数据的采集和存储变得更为便利和快捷，我们生活的世界每天产生的数据呈爆炸式增长。美国互联网数据中心的资料显示，每年在互联网上的数据以 50%的比例增长。每时每刻，海量数据都在源源不断地产生。

2011 年 2 月，《Science》杂志在社论中指出，“数据推动着科学的发展”[1]。2013 年 3 月 5 日，出席全国两会的人大代表、安徽移动总经理郑杰建议将“发展大数据”上升到国家战略。他认为，发展大数据技术的关键并不仅仅是对海量数据的掌握，最重要的是如何专业化地处理这些有意义的数据。2013 年 3 月 29 日倪光南院士在武汉大学“云计算与软件服务工程创新发展高峰论坛”上作《迎接大数据时代的来临》的报告，他认为，大量产生的数据加上云计算的发展，为大数据提供了合适的环境和处理能力，推动了数据挖掘、商业智能向大数据发展。数据挖掘就是为解决这一问题而产生的研究领域，它是从存放在数据库、数据仓库或其他信息库的大量数据中“挖掘”有趣知识的过程[2]。

在数据挖掘中，描述数据的特征维数越来越高，然而其中大部分特征可能和挖掘任务不相关或特征之间存在相互冗余，使得数据挖掘中学习算法的时空复杂度增高、效果变差，这种现象被称为“维数灾难”。面对“维数灾难”，如何降低维数显得非常迫切，特征选择就是一种有效的降维方法。通过特征选择，消除数据中的无关和冗余特征，不仅可以提高从大量数据中发现知识的效率，而且能够改善后期得到的分类器性能。因此，特征选择成为数据挖掘中的重要研究分支。

现实世界中的数据纷繁复杂，不可避免地存在大量的噪声、不相关和不一致性，因此，对特征选择的要求不断提高。粗糙集（Rough Set，RS）[3]理论是波兰科学院 Z.Pawlak 院士于 1982 年提出的，是一种相对较新的软计算工具，能够处理不确定和不精确信息。它在特征选择算法中得到广泛应用，已逐渐成为一种重要的特征选择理论框架。基于粗糙集的特征选择，要求最终得到的特征子集，不仅其分类能力与原始特征集合的分类能力一致，而且具有最少的基数。

自 Z.Pawlark 教授提出粗糙集理论以来，经过短短 30 余年的发展，涌现出大

量基于粗糙集的特征选择算法，根据其采用搜索方法的不同，可分为三大类：穷举法、启发式方法和随机方法。穷举法[4-7]，是指首先求出所有满足要求的特征子集，然后从中选取具有最少基数的特征子集。很明显，这种解决问题的方法并不适合于大规模数据集。已经有文献证明，求出所有满足要求的特征子集是一个NP-难问题[8]。因此，就必须考虑启发式方法。启发式方法[9-17]从一个特定的特征集合（空集或全集）出发，使用启发式信息来引导特征选择过程，不断添加或删除特征，直至得到满足要求的特征子集。如果数据中含有的噪声和特征数目不多，那么启发式方法效果较好，能得到较优的特征子集，但无法确保得到最优特征子集。随机方法[18-23]主要利用遗传算法等随机算法健壮的搜索能力来产生最优特征子集，虽然能够提供一个更好的特征选择解决方案，但是操作比较耗时，而且也无法保证每次都能得到最优特征子集。

综上所述，这三类方法中能够确保得到最优特征子集的只有穷举法，但穷举法需要求出所有满足要求的特征子集，计算复杂度高，需要消耗大量时间，所以不适合处理大数据集；启发式方法和随机方法操作简单，运行速度较快，但却无法保证得到最优特征子集。因此，探索更有效的特征选择算法势在必行。

群智能（Swarm Intelligence，SI）[24]是指无智能或具有简单智能的个体组织在一起，如蚁群、鸟群和蜂群等，通过相互之间的协作而表现出智能行为的特性。群智能方法是近年发展起来的新型仿生智能优化算法，受到研究者的广泛关注，已经成为人工智能、数据挖掘、社会经济以及生物等交叉学科的研究热点。

群智能中的代表性算法如：蚁群优化（Ant Colony Optimization，ACO）算法[25]、粒子群优化（Particle Swarm Optimization，PSO）算法[26]和人工蜂群（Artificial Bee Colony，ABC）算法[27]等，自 20 世纪末提出以来，已经广泛应用于人工智能、数据挖掘和工业生产等领域。大量文献证明其能够解决不同领域的问题，特别在解决许多问题时表现出比传统优化算法更好的性能。群智能方法和人类社会经济生活紧密相关，拥有广阔的市场前景，无论是从理论研究还是应用研究的角度，对群智能方法进行研究都具有重要的学术意义和现实价值[28]。

在基于粗糙集的特征选择过程中，群智能方法已经崭露头角，并彰显出独特的优势。然而，群智能方法还有一些主要问题需要解决：早熟问题广泛存在不容忽视；对于大规模优化问题，算法后期容易出现停滞；参数多凭经验设置，对具体问题和应用环境依赖性大。

本书在群智能的代表性算法中，选取 ACO 算法、PSO 算法和 ABC 算法，深入研究它们在基于粗糙集的特征选择中的应用。原因在于，ACO 算法擅长处理组合优化问题，而特征选择本质上就是一个组合优化问题；PSO 算法在离散空间的优化方面较为成熟；ABC 算法的研究始于 2005 年，才刚刚起步，方兴未艾，发

展空间很大。

因此，本书以粗糙集特征选择为基础，重点研究群智能方法及其在粗糙集特征选择中的应用。充分发挥群智能方法的寻优能力来搜索最优特征子集，同时利用粗糙集的计算能力来评价特征子集的优劣，两者有机结合，优势互补，促使特征选择朝着特征子集分类能力最强、基数最少的方向前进，不断逼近全局最优解。群智能方法在粗糙集特征选择中的应用，为特征选择的研究提供了一种有效的解决方案，注入了新的活力，顺应了时代发展的要求，理论意义和应用价值都非常巨大。

1.2 特征选择概述

特征选择是根据特定的评价标准从原始特征集合中选择一部分特征构成一个特征子集，该特征子集能够保持原始特征集合的分类能力，同时只包含原始特征集中最少的特征[29]。通过特征选择，删除原始特征集合中大量的无关和冗余特征，不仅可以降维，解决“维数灾难”问题，而且选择后的结果更易于理解。

根据是否在数据样本中包含分类标签，特征选择可以分为三种类型：有监督特征选择、无监督特征选择和半监督特征选择。有监督特征选择是指数据样本中包含分类标签，而且该分类信息将用以指导整个特征选择过程。目前，有监督特征选择已经成为特征选择领域的主流研究方向，其得到的特征子集不仅分类能力强，而且包含的特征数目少。无监督特征选择是指数据样本中不包含分类标签，整个特征选择过程仅利用数据本身具有的内在关系，通过一些特征评价指标来进行特征选择。半监督特征选择介于两者之间，其数据样本中既有少量包含分类标签的样本(称为已标记样本)，又有大量不包含分类标签的样本(称为未标记样本)。首先对已标记样本集采用有监督特征选择，利用其分类信息指导特征选择过程得到特征子集，然后结合未标记样本集对该特征子集作进一步地选择或评价。本书研究有监督特征选择问题。

特征选择的本质是一个组合优化问题。从大小为 n 的特征集合中选择一个最优特征子集，其搜索空间可达 2^n-1。因此，特征选择中采取何种搜索策略是非常重要的。用于特征子集搜索的主要策略如下：

（1）全局最优搜索策略：包括穷举法和分支定界法。穷举法可以搜索到所有的特征子集，但计算量大，尤其特征数较多时几乎不可行。Narendra 和 Fukunaga[30] 提出的分支定界法以及 Chen[31]提出的改进方法，均通过剪枝策略减少计算量，而且其具有回溯功能，可以涵盖所有的特征组合，但算法复杂性仍然较高，并且要求评价函数具有单调性。

（2）启发式（或序列）搜索策略：在搜索过程中，将特征依据一定的次序，不断向当前特征子集进行添加或剔除，直至得到优化特征子集。比较典型的有Whitney[32]提出的前向搜索，Marill 和 Green[33]提出的后向搜索，Stearnsl[34]提出向前加 l 个特征和向后减 r 个特征进行前后相结合的浮动搜索等等。启发式搜索较容易实现，计算复杂度相对较小，但容易陷入局部最优。

（3）随机搜索策略：首先随机产生一些候选特征子集，然后依照一定的启发式信息和规则不断对其更新，直至逐步逼近全局最优解。例如：禁忌搜索[35]、模拟退火法[36]和遗传算法[37]等等。随机搜索策略计算量大，所需时间长。

上述三种搜索策略各有所长，也各有所短，需要在实际应用时，根据具体情况进行选择。全局最优搜索策略适合特征数目较少的数据集；启发式搜索策略速度快，但不一定能够得到最优特征子集；随机搜索策略介于两者之间。

特征选择方法依据是否独立于后续学习算法，可分为 Filter 方法[38]、Wrapper 方法[39]和 Embedded 方法[40]三类，其中 Filter 方法和 Wrapper 方法最常用。Filter 方法独立于后续学习算法，仅按照特征的重要性来构造特征子集，其关键是特征重要性的定义。常用的特征重要性计算方法有卡方检验、信息增益、基尼系数等[41]。Filter 方法需要一个阈值作为特征选择的停止准则。该方法的特点是速度快，但当所选特征与后续学习算法紧密相关时，偏差较大。经典的 Relief 特征选择算法[42]就是 Filter 方法。

Wrapper 方法依赖于后续学习算法，需要将训练样本分成训练子集和测试子集两部分，并根据后续学习算法的训练准确率来评价特征子集的性能。因此，Wrapper 算法偏差小，但对后续学习算法依赖大，并且所需计算量较大。

Embedded 方法也依赖于后续学习算法，其将特征选择过程嵌入到学习算法训练分类器的过程中，通过一个优化函数模型实现特征选择。该方法的特点是不需要将训练样本分成训练子集和测试子集，后期也不必训练分类器来对特征子集进行评估，因此速度快、效率高，但优化函数模型的构造比较困难。

近年来，学者们倾向于采用混合特征选择方法来选择最优特征子集[43]，这也是目前特征选择方法研究的一个新趋势。最为常用的是将 Filter 和 Wrapper 相结合来选择特征子集，首先使用 Filter 方法将原始特征集中的无关和冗余特征进行过滤，然后在此基础上使用 Wrapper 寻找最优特征子集。Filter 和 Wrapper 方法的结合，优势互补，可以提高特征选择的性能并降低时间复杂度。2007 年，Uncu 和 Turksen[43]利用函数依赖概念、相关系数和 K-近邻来实现特征的过滤和封装。2008 年，王树林等[44]以肿瘤样本集的分类性能作为启发式反馈信息，基于支持向量机提出一种 Filter-Wrapper 混合方法进行特征选择。2011 年，Akadi 等[45]结合最小冗余最大相关算法和遗传算法提出一种 Filter-Wrapper 混合方法选择特征子集。

2012 年，Foithong 等[46]基于互信息和粗糙集提出一种 Filter-Wrapper 混合方法选择特征子集，首先利用互信息取代用户定义的参数来过滤候选特征，然后采用 Wrapper 方法搜索候选特征集空间，可以降低计算成本，避免陷入局部最优。

1.3 国内外研究现状

1.3.1 基于粗糙集的特征选择研究进展

粗糙集理论是一种处理模糊和不精确问题的新型数学工具[3]。最初关于粗糙集理论的研究主要集中在东欧国家，当时并没有引起重视。1991 年，粗糙集理论创始人 Z.Pawlak 出版了他的第一本粗糙集专著，标志着粗糙集理论与应用的研究进入了活跃时期。国际人工智能与模式识别的研究学者开始广泛关注粗糙集理论的应用研究，特别是在数据挖掘、决策分析、模式识别、机器学习和智能控制等领域。为了给广大研究人员提供学术交流的机会，从 1992 年开始，每年举办一届粗糙集理论的国际学术会议，自此，关于粗糙集理论的文献如雨后春笋不断涌现。

基于粗糙集的特征选择，在粗糙集理论中称作属性约简，它是粗糙集理论的一个重要研究课题。所谓属性约简就是在保持属性集合分类能力不变的前提下，删除其中冗余的属性。因此，粗糙集理论已经广泛应用于构造特征选择算法。故在本书中，对于属性或特征、基于粗糙集的特征选择或粗糙集特征选择或属性约简，就不再进行区分。

自粗糙集理论提出以来，短短 30 余年里，涌现出了大量基于粗糙集的特征选择算法，根据其采用搜索方法的不同，可分为三大类：穷举法、启发式方法和随机方法。

穷举法，是指首先求出所有满足要求的特征子集，然后从中选取基数最少的特征子集。区分矩阵（差别矩阵）是粗糙集理论的核心概念之一。1992 年，Skowron 和 Rauszer[4]首先提出区分矩阵的概念，然后基于此提出求解信息系统完备（所有）约简的方法。利用任意两个对象之间的不同特征，来描述数据集中蕴涵的分类知识，然后从这些数据中构造出区分函数，最后转化成最简形式。为了加快计算速度，1999 年和 2000 年，Starzyk 等[5,6]使用强等价关系来简化区分函数。2009 年，Yao 和 Zhao[7]利用经典高斯消去法对区分矩阵进行简化，通过矩阵运算直接得到约简。总之，尽管穷举法在理论上很完备，可以得到信息系统的所有约简结果，但仍然避免不了“组合爆炸”这一难题。已经证明，求解所有约简是 NP-难问题[8]。因此，必须考虑启发式方法。

启发式方法是一种近似算法，实现过程简单、快速，实际应用非常广泛。启

发式方法中，通常采用启发式信息来引导特征选择过程，可以从一个空特征集或特征核开始，然后根据启发式信息不断添加特征直至得到满足要求的特征子集，也称为前向选择法；或者从特征全集开始，根据启发式信息不断删除特征直至得到满足要求的特征子集，也称为后向删除法。启发式信息可采用粗糙集的特征重要性来定义，各种启发式方法的根本区别就在于对特征重要性的定义不同。1995年，Hu 和 Cereone[9]提出基于正域的启发式特征选择算法。首先以去掉特征后正域的变化大小来定义特征重要性，然后从特征核出发，按照特征重要性的大小由大到小逐个加入特征，直至特征子集的依赖度与原始特征集的依赖度一致；接着用向后删除的方法，逐个检查所得结果中的每个特征，凡是删除后不影响特征子集依赖度的特征，均为冗余特征，最后得到的特征子集就是最优特征子集。后来，Chouchoulas 和 Shen[10]在此基础上作出一些改进。随着粗糙集理论的发展，信息熵被广泛应用于度量信息系统的不确定性。2002 年和 2003 年，王国胤等[11,12]给出决策表特征核的计算方法，并提出基于条件熵的特征选择算法。首先基于 Shannon 条件熵，将添加特征后条件熵的变化大小定义为特征重要性，然后以特征核为出发点，按照特征重要性从大到小逐个加入特征，直到特征子集相对于决策特征的条件熵与原始特征集相对于决策特征的条件熵相等为止。2002 年，Liang 等[13]针对 Shannon 熵无法度量粗糙集的模糊性，引入互补熵，并设计出基于互补熵的启发式特征选择算法。此外，2008 年，Qian 和 Liang[14]提出组合熵的概念。1999 年，苗夺谦和胡桂荣[15]分析不同特征之间的互信息，利用互信息的变化大小来定义特征重要性，提出一种基于互信息的启发式特征选择算法。2003 年，Hu 等[16]将区分矩阵中特征的出现频率作为特征重要性，给出一种快速的特征排序机制，并在此基础上提出基于特征频率的特征选择算法。2010 年，为了加快启发式方法的计算效率，Qian 等[17]提出一种正向近似的理论框架，用于加速特征选择的启发式过程，并证明其可行性和高效性。对于上述启发式方法，由于不存在完备的启发式信息，使用特征重要性来选择下一个特征会导致搜索沿着一条非最优的途径进行，无法保证最终结果的最优性，因此，启发式方法并不能保证找到最优特征子集。

随机方法是一种相对较新的方法。1995 年，Wroblewski[18]提出三种遗传算法（Genetic Algorithm，GA）来产生最小约简。第一种算法是经典 GA 算法，个体采用二进制位串表示，算法速度很快，但有时会陷入局部最优；后两种算法是基于置换编码和贪婪算法，能够得到更好的结果，但需要增加计算时间。2002 年，Zhai 等[19]提出一个集成的特征提取方法，并在此基础上建立特征提取原型系统。该系统成功地将粗糙集处理不确定性的能力和 GA 算法健壮的搜索能力进行集成，然后用于简化产品质量评价。后来，又有一些学者在 GA 算法的基础上，引

入其他随机算法。2007 年，陈友等[20]将 GA 算法和禁忌搜索算法进行混合用于特征选择，构建轻量级入侵检测系统。2008 年，Hedar 等[21]提出一种基于内存的启发式禁忌搜索算法用于特征选择，可以节约计算成本。2009 年，张昊等[22]在自适应 GA 算法中加入模拟退火的思想，进行特征选择，可以加速算法收敛，避免陷入局部最优，提高特征选择的效率。2011 年，Abdullah 等[23]提出一个再热模拟退火算法用于特征选择，再热可以帮助算法更好地探索搜索空间，找到更好的解，从而逃离局部最优。上述这些随机方法，虽然能够提供一个更好的特征选择解决方案，但是操作非常耗时，需要进行大量的计算，而且也无法保证每次都能得到最优特征子集。

综上所述，这三类方法中能够确保得到最优特征子集的只有穷举法，但穷举法需要求出所有满足要求的特征子集，计算复杂度高，并且需要消耗大量时间，所以不适合处理大数据集；启发式方法，简单、快速且效率较高，但由于不存在完备的启发式信息，并不能保证找到最优特征子集；随机方法，虽然能够提供一个更好的特征选择解决方案，但是操作非常耗时，需要进行大量计算，而且也无法保证每次都能得到最优特征子集。因此，探索更有效的特征选择算法势在必行。

在基于粗糙集的特征选择过程中，已有一些群智能方法不断引入进来。2003 年，Jensen 和 Shen[47]采用 ACO 算法用于特征选择；2007 年，Wang 等[48]基于粗糙集和 PSO 算法，提出一种新的特征选择策略；2010 年，Bae 等[49]受 Wang 的启发，提出一种新算法，即智能动态群（Intelligent Dynamic Swarm，IDS）。这是一个改进的 PSO 算法、粗糙集和 *K*-均值混合方法，首先采用 *K*-均值聚类算法处理连续变量，然后使用 IDS 算法进行特征选择。下面介绍群智能的研究进展。

1.3.2 群智能研究进展

群智能的概念源于 20 世纪 80 年代，人们对社会性动物（如蚁群、鸟群、蜂群等）的自组织行为发生了浓厚兴趣。研究人员发现：虽然它们单一个体的智能不高，也没有集中指挥，但它们组成的群体却能够协同工作，建立巢穴、集中食物、哺育后代等，发挥超出个体的智能。表 1.1 简单总结了群智能的发展历程。

从表 1.1 中可以看到，虽然群智能发展时间不长，但已受到研究者的广泛关注，成为人工智能、数据挖掘、社会经济以及生物等交叉学科的研究热点。

群智能中的代表性算法如 ACO 算法、PSO 算法和 ABC 算法等，都属于启发式随机搜索算法，它们依靠群体之间的信息共享来求解复杂问题，体现了群智能的协作性、简单性和分布性等特点。群智能为许多传统方法较难解决的组合优化、知识发现和 NP-难问题提供新的求解方案，为许多前瞻性研究提供新的思路，具有重要的学术意义和现实价值。

表 1.1 群智能发展历程

时间	大事记
1991 年	Colorni 等[25]提出 ACO 算法
1995 年	Kennedy 和 Eberhart [26]提出 PSO 算法
1998 年	Dorigo 等组织两年一次的关于 ACO 算法和群智能的国际会议
1999 年	国际进化计算大会召开 ACO 算法专题会议
1999 年	Bonabeau 等[24]编写群智能的专著，提出群智能的概念
2002 年	IEEE 进化计算汇刊出版 ACO 算法和群智能的专辑
2005 年	Karaboga[27]提出 ABC 算法

由于 ACO 算法适合于求解组合优化问题，PSO 算法在离散空间优化方面较为成熟，ABC 算法处于起步阶段，发展空间很大，所以本书围绕这三种算法开展研究。

1.3.2.1 蚁群优化算法研究进展

ACO 算法是一种仿生智能优化算法，模拟昆虫王国中蚂蚁群体的觅食行为。1991 年，Colorni 等[25]首次提出 ACO 算法，但直到 1996 年才引起国际学术界的关注，事情发展的契机是 Dorigo 等[50]发表的一篇文章 *Ant System: Optimization by A Colony of Cooperating Agents*。而 ACO 算法的第一部专著 *Ant Colony Optimization*，是 Dorigo 和 Stutzle[51]于 2004 年出版的，内容详实、系统、权威，成为研究人员的经典参考资料。

ACO 算法的蓬勃发展，为许多寻优问题提供一种新的解决方案。众多学者致力于 ACO 算法的研究，主要体现在三个方面：

（1）ACO 算法的理论研究。2000 年，Gutjahr[52]首次对 ACO 算法的收敛性进行证明，虽然是在一些假设前提下，但仍具有重要意义。2002 年，Stutzle 和 Dorigo[53]提出一种简化的 ACO 算法，认为该算法对具有组合优化性质的极小化问题总能找到全局最优解。2004 年，Badr 和 Fahmy[54]从分支随机路径和分支过程的角度，研究 ACO 算法的收敛性。

2003 年，孙焘等[55]将 GA 算法与 ACO 算法进行融合，并从 Markov 随机过程的角度，分析该混合算法的收敛性。2009 年，苏兆品等[56]首先把旅行商问题（Traveling Salesman Problem，TSP）描述为一类 ACO 算法的数学模型，然后分解状态空间，构筑反射壁，最后从鞅理论的角度，证明该类 ACO 算法不仅具有几乎处处强收敛性，而且能够在有限步内快速收敛，得到全局最优解。

（2）ACO 算法的改进。1996 年，Dorigo 和 Gambardella[57]在原有蚂蚁系统基础上，结合强化学习提出 Ant-Q 系统。重点研究 Ant-Q 对参数的敏感性，并调

查蚂蚁之间的协同效应。1997 年，Stutzle 和 Hoos[58]提出 Max-Min 蚂蚁系统。该系统仅对本次遍历中最优路径上的信息素进行增加，并将其值限定在一定范围之内，对信息素更新机制进行改进，使得每条路径上的信息素存在较大的浓度差异，从而加快收敛，避免陷入局部最优。1999 年，Bullnheimer 等[59]提出一种基于排序的新蚂蚁系统。当所有蚂蚁结束一次遍历后，首先按照蚂蚁所走路径的长度进行升序排列，然后按照每个解的质量给予权重，最后根据解的不同权重来更新信息素。很多学者针对 ACO 算法本身具有的并行性开展研究，提出许多改进措施。2011 年，Pedemonte 等[60]对并行 ACO 算法的研究进行综述，介绍并行计算技术在 ACO 算法中的应用情况。

1999 年，吴庆洪等[61]提出一种具有变异特征的 ACO 算法。该算法引入变异算子，利用逆转变异方式，改善蚁群的性能，减少计算时间，加快算法收敛。2002 年，王颖和谢剑英等[62]提出一种自适应改变信息素挥发系数的 ACO 算法，在收敛速度不受影响的前提下，提高解的全局优化性能。2003 年，熊伟清等[63]将遗传变异算子引入 ACO 算法。通过设置信息阈值修改选择策略，让蚂蚁在初始时刻有较多选择，增加多样性；同时，对路径选择策略进行改进，全局修正信息素更新规则；引入变异，通过逆转变异和插入变异产生新解，进行局部优化，增加个体多样性，改善整个群体性能；并且对蚁群中蚂蚁进行分工，减少每只蚂蚁的搜索空间，增强算法整体搜索能力。2007 年，陈峻和章春芳[64]提出一种采用自适应信息交换策略的并行 ACO 算法。首先对处理机之间的信息交流提出两种策略，然后衡量优化过程中信息素在各路径上的分布均匀度，对信息素更新策略进行自适应地调节，可以有效地缓解快速收敛和早熟停滞现象之间的矛盾。

（3）ACO 算法与其他算法的融合及应用。很多学者研究 ACO 算法与其他算法的融合，如：GA 算法[65-71]，模拟退火算法[72-76]，免疫算法[77-82]，神经网络[83-87]等，更有学者将 ACO 算法与多个算法进行大融合[88-90]。多种算法的融合，可以实现优势互补，提高收敛速度，改善算法性能。

2007 年，Jangam 和 Chakraborti[65]将 GA 算法和 ACO 算法进行混合，用于核酸序列的两两比对。该混合算法首先采用 ACO 算法获取一个核酸序列，然后再利用精英 GA 算法，使用原始的选择算子，结合一个新的多点交叉变异算子，生成一个核酸的精确序列，最后进行两个核酸序列的比对。2008 年，Lee 等[66]则将 GA 算法与 ACO 算法用于多序列的比对，其中 GA 算法提供多样性，ACO 算法负责跳出局部最优。2011 年，Chen 和 Chien[67]采用并行遗传蚁群系统来解决 TSP 问题。2012 年，Ciornei 和 Kyriakides[68]提出一种包含特殊连续域 ACO 算法和 GA 算法的混合算法，并证明其收敛性。2004 年，邵晓魏等[69]采用 GA 算法生成信息素分布，首先均匀分割问题空间，然后采用 GA 算法将初始种群均匀分散在解空

间，并利用 ACO 算法求出精确解，两种算法优势互补，防止过早收敛，加快收敛速度。同年，朱庆保和杨志军[70]提出一种高速收敛算法。该算法对信息素采用一种新颖的动态更新策略，让所有蚂蚁在每一次搜索过程中都发挥出最大贡献；同时，每次搜索结束后，对搜索的结果引入独特变异来进行优化，可以大幅度提升算法的收敛速度。2009 年，肖宏峰和谭冠政[71]将 GA 算法融入 ACO 算法，提出两种新策略：一种是先利用 GA 算法找到一组解，然后再用 ACO 算法寻找最优解；另一种是利用 GA 算法，采用交叉操作，产生 ACO 算法的新旅行路径。

2006 年，傅鹏等[72]提出一种新的 QoS 路由发现方法，将 ACO 算法与模拟退火算法进行结合，针对可用 QoS 路由，利用 ACO 算法增加对其发现的概率；同时利用模拟退火算法调整 ACO 算法的搜索方向，减少停滞现象的发生。2008 年，Musa 和 Chen[73]利用几种算法的组合来解决动态吞吐量最大化问题，包括一个简单的贪婪排序算法、两个模拟退火算法和两个 ACO 算法。2009 年，刘波和蒙培生[74]采用模拟退火算法使信息素分布集中，加快收敛，并结合 3opt 局部优化算法提高效率，同时证明该算法收敛。2012 年，Niksirat 等[75]利用一个贪婪模拟退火算法和一个双种群的双向搜索蚁群系统来解决运输网络中包含多个点的 K-最短可行路径问题。同年，张亚明等[76]提出一种适用于多跳 WSNs 的基于蚁群模拟退火算法的移动 Agent 访问路径规划模型。

2006 年，钟一文和杨建刚[77]提出一种免疫 ACO 算法，采用 ACO 算法对任务调度的优先队列进行进化，并使用免疫原理保持蚁群多样性，避免早熟停滞。2009 年，闭应洲等[78]提出基于免疫修复的 ACO 算法。采用免疫原理识别候选解中的“病变”成分，并对其进行修复，提高候选解的质量，加快正反馈过程。2010 年，刘朝华等[79]提出双态免疫 ACO 算法，一方面将蚂蚁划分成两种状态，使解的搜索空间扩大，早熟停滞现象得到有效的抑制；另一方面采用几种免疫算子进行运算，从而得到精英蚂蚁，并将局部最优免疫策略引入抗体记忆库，不仅可以快速收敛，而且能够提高求解的精度。2011 年，万芳等[80]提出基于免疫进化的 ACO 算法，不仅利用免疫算法的快速收敛优势，而且在 ACO 算法中增加扰动策略，有效地克服 ACO 算法存在的问题，并在滦河下游六水库联合供水优化调度中进行应用。同年，Huang 和 Cen[81]基于环境建模，结合 ACO 算法和免疫调节，提出一种新的全局路径规划算法。Wang 等[82]将免疫算法与 ACO 算法结合寻找最佳飞行路线。该算法首先在飞行区域随机生成初始路线，然后用克隆选择算法搜索好的路线，得到一组风险和耗油成本最小的路线；同时，在这些路线附近放置一些初始信息素，在此基础上，再使用 ACO 算法搜索风险和耗油成本最小的最优路线。