

信息科学技术学术著作丛书

参数计算导论

王建新 冯启龙 著



科学出版社

信息科学技术学术著作丛书

参数计算导论

王建新 冯启龙 著

科学出版社

北 京

内 容 简 介

本书较全面地介绍了参数计算理论的提出背景、理论范畴、相关算法设计与分析技术以及参数计算的实际应用。具体阐述了核心化技术、局部贪婪、递归压缩、分支搜索、随机方法、彩色编码、固定参数枚举技术,并从生物信息计算、计算机网络角度探讨了参数计算理论的实际工程应用价值。

本书设置了理论深度不同的章节,以适应不同层次的读者对象,既可以作为计算机算法领域研究生的教材,也适用于计算机科学领域的科研人员。

图书在版编目(CIP)数据

参数计算导论/王建新,冯启龙著. —北京:科学出版社,2014

(信息科学技术学术著作丛书)

ISBN 978-7-03-036899-7

I. 参… II. ①王…②冯… III. 参数-计算-应用-信息学 IV. G201

中国版本图书馆 CIP 数据核字(2013)第 044095 号

责任编辑:魏英杰 杨向萍 / 责任校对:胡小洁

责任印制:肖 兴 / 封面设计:陈 敬

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>

文林印刷厂 印刷

科学出版社发行 各地新华书店经销



2014 年 6 月第 一 版 开本:B5(720×1000)

2014 年 6 月第一次印刷 印张:15 1/2

字数:312 000

定价:80.00 元

(如有印装质量问题,我社负责调换)

《信息科学技术学术著作丛书》序

21 世纪是信息科学技术发生深刻变革的时代,一场以网络科学、高性能计算和仿真、智能科学、计算思维为特征的信息科学革命正在兴起。信息科学技术正在逐步融入各个应用领域并与生物、纳米、认知等交织在一起,悄然改变着我们的生活方式。信息科学技术已经成为人类社会进步过程中发展最快、交叉渗透性最强、应用面最广的关键技术。

如何进一步推动我国信息科学技术的研究与发展;如何将信息技术发展的新理论、新方法与研究成果转化为社会发展的新动力;如何抓住信息技术深刻发展变革的机遇,提升我国自主创新和可持续发展的能力?这些问题的解答都离不开我国科技工作者和工程技术人员的求索和艰辛付出。为这些科技工作者和工程技术人员提供一个良好的出版环境和平台,将这些科技成就迅速转化为智力成果,将对我国信息科学技术的发展起到重要的推动作用。

《信息科学技术学术著作丛书》是科学出版社在广泛征求专家意见的基础上,经过长期考察、反复论证之后组织出版的。这套丛书旨在传播网络科学和未来网络技术,微电子、光电子和量子信息技术、超级计算机、软件和信息存储技术,数据知识化和基于知识处理的未来信息服务业,低成本信息化和用信息技术提升传统产业,智能与认知科学、生物信息学、社会信息学等前沿交叉科学,信息科学基础理论,信息安全等几个未来信息科学技术重点发展领域的优秀科研成果。丛书力争起点高、内容新、导向性强,具有一定的原创性;体现出科学出版社“高层次、高质量、高水平”的特色和“严肃、严密、严格”的优良作风。

希望这套丛书的出版,能为我国信息科学技术的发展、创新和突破带来一些启迪和帮助。同时,欢迎广大读者提出好的建议,以促进和完善丛书的出版工作。

中国工程院院士

原中国科学院计算技术研究所所长



前 言

20 世纪 90 年代末, Fellows、Downey 和 Chen 等提出并完善了参数计算理论、参数算法设计与分析技术, 旨在利用参数计算理论和技术设计难解问题的有效实用算法。参数计算理论的提出源于人们对难解计算问题进行的详细结构性分析。这些分析结果, 在实际工程应用中的发现难解问题可以通过引入一个或多个参数来有效求解, 更重要的是, 这些参数的值一般比较小, 充分利用小参数这一特征进行数学建模, 不少实际工程应用中难解计算问题就变得实际可解了。例如, 生物信息计算领域基因序列冲突问题可转化成经典 NP 难题(点覆盖)。而在实际生物应用背景中, 问题点覆盖的规模 k 不超过 60。基于这一参数特性, 人们提出时间复杂度为 $O(1.285^k + kn)$ 的参数算法, 其中 n 是给定图的大小。这一算法使得基因序列冲突问题可在较短的时间内得到解决。

自 20 世纪 90 年代提出以来, 参数计算理论引起了计算机理论领域的广泛关注。近年来更是以成为计算机理论和应用领域的研究热点。当前许多国家, 如美国、德国、澳大利亚、挪威、加拿大等, 积极投入力量进行该领域的研究。有关这方面研究主题的论文已大量发表在各种重要的计算机科学与技术的高档次期刊和会议上。关于参数计算算法和理论的特刊也陆续在著名的国际期刊, 如 *Journal of Computer and System Sciences*、*Theoretical Computer Science*、*The Computer Journal*、*Algorithmica*、*Theory of Computing Systems* 上发表。更值得注意的是, 近年来大量与参数计算直接或间接相关的研究成果已发表在理论计算机科学领域外的期刊和会议上, 如《人工智能》、《计算生物化学》、《数据库系统》、《机器人》和《软件工程》等。这都表明近年来参数计算理论的研究成果不仅在计算理论上取得了很大的突破和进展, 而且在实际工程应用领域得到了应用, 参数计算的思想已经在解决工程难解问题中得到认可, 并且越来越多地被应用到工程难解问题的建模和算法设计中。

参数计算理论是求解难解问题的一种有效途径。尽管近年来参数计算理论在国内外的研究已经成为热点, 但是参数计算理论正处于一个发展阶段, 还需要进一步丰富和发展。另外, 参数计算理论领域还存在诸多挑战, 如新的算法设计与分析技术、核心化下界理论、参数算法工程实现、应用领域的拓宽等。本书试图从参数计算理论的基本思想、方法和技术, 在工程难解问题求解中的应用等多个方面, 结合近年参数计算理论的研究成果和面临的挑战介绍参数计算理论。

我们一直致力于参数计算、生物信息计算等方面的研究。沿着参数计算的国

际前沿研究方向,凭借在参数计算理论、参数算法设计与分析技术、固定参数枚举技术、参数计算实际应用等方面的研究基础,我们取得了一系列的研究成果。本书相关的研究得到了国家 973 计划前期研究专项项目“信息处理中参数计算算法研究与实现”的资助。基于该项目的资助,课题组以现实世界信息处理中出现的难解问题为出发点,建立了相应难解问题的参数化计算模型,研究了真实有效的算法,并通过设计与实现相应问题的计算机软件进行难解问题求解方法的验证与真实求解。其研究成果为系统解决信息处理中的关键与热点问题提供了新的理论基础和研究方法。此外,课题组还得到了国家自然科学基金项目“难解问题的核心化技术及其应用研究”、“参数计算理论及应用”、高等学校博士学科点专项科研项目“彩色编码技术及应用研究”、教育部创新团队项目“计算机优化算法及其应用技术研究”等资助。本书相关的研究还获得国家自然科学基金重点项目“生物信息学中的相关组合理论和算法研究”的资助,基于在生物信息学中的研究基础,将参数计算理论思想运用到生物信息学中若干难解问题的求解中,如单体型组装问题、基序查找问题等,得到了一系列有效的参数算法,提高了求解效率,相关研究成果发表在国际著名期刊和会议上。

全书共 12 章。第 1 章介绍参数计算的提出背景,并从参数化顶点覆盖、二分图受约束点覆盖、程序类型检测、单体型组装等具体应用实例出发阐述参数计算方法的优劣及其发展现状。第 2 章从 NP 完全理论、参数计算的基本理论框架、固定参数可解理论、固定参数不可解理论、参数化方法等方面进行介绍。第 3 章介绍较为流行的核心化技术及其具体应用,如 NT 定理、皇冠分解、极值归纳、随机方法等,并概述了核心化下界理论。第 4 章以参数化点覆盖问题、参数化顶点反馈集问题、超平面覆盖问题为例,介绍常规的分支搜索法、基于隐含参数的分支搜索法和基于组合的分支搜索法等分支技术。第 5 章首先介绍递归压缩技术的提出背景、技术原理,并以参数化顶点反馈集问题为例阐述递归压缩技术的具体应用,并以 3-Set Packing 问题为例介绍局部贪婪技术的原理和应用,最后介绍递归压缩技术与局部贪婪技术的综合运用。第 6 章以 k -Path 问题、3-Set Packing 问题为例介绍随机方法在参数算法设计与分析中的应用,并介绍较为流行的确定化方法 (n, k) -Universal Set 及其具体应用。第 7 章介绍彩色编码技术原理及其应用,主要分析随机彩色编码方法、确定化彩色编码方法和 $n \leq 2k$ 情况下的彩色编码技术,并分别以 k -Path 问题、3-Set Packing 问题、生物信息计算中 Motif Finding 问题为例介绍各类彩色编码技术的具体应用。第 8 章介绍关于平面图的相关概念,并介绍平面图算法设计技术层状分割及其在平面支配集问题中的应用、平面图核心化技术及其应用。第 9 章具体介绍固定参数枚举理论,并分别以加权顶点覆盖问题、加权 3D-Matching 问题、加权顶点反馈集问题为例阐述基于分支搜索的枚举技术、基于彩色编码的枚举技术、基于递归压缩的枚举技术。第 10 章介绍参数计算与近似算

法的关系,具体介绍广义参数化近似算法、标准参数化近似算法,并以 3-D Matching计数问题为例介绍参数化随机近似算法,最后介绍不存在参数化近似算法的问题。第 11 章介绍树分解技术相关定义,基于树分解的参数算法设计及其应用。第 12 章从若干生物信息计算问题、计算机网络问题出发介绍参数计算理论的应用。附录给出关于计算模型、计算复杂性、集合论、图论等方面的基本定义和符号。

借本书出版之机,作者向资助、支持和关心过相关研究工作及本书出版的各有关单位和个人致以诚挚的谢意!感谢科技部、自然科学基金委的资助。感谢国内外学者与本课题组在参数计算和生物信息计算等方面的合作与交流,特别是美国德州农机大学陈建二教授、澳大利亚纽卡斯尔大学 Fellows 教授、美国乔治亚州立大学潘毅教授、德国 Saarland 大学郭昊教授、美国佐治亚大学蔡黎明教授、北京航空航天大学大学的许可教授等。最后要感谢作者的同事及研究生们,他们参与了相关课题的研究工作,付出了辛勤劳动,本书的出版是大家共同努力的结果。

限于作者水平,书中不妥之处在所难免,恳请广大读者批评指正。

作者

目 录

《信息科学技术学术著作丛书》序

前言

第 1 章 导引	1
第 2 章 参数计算简介	5
2.1 NP 完全理论	5
2.2 固定参数可解	7
2.3 固定参数不可解	8
2.4 固定参数枚举	10
2.5 参数化方法	10
2.6 本章小结	12
第 3 章 核心化	13
3.1 NT 定理	15
3.1.1 基于最大匹配的 NT 算法	15
3.1.2 基于线性规划的 NT 算法	18
3.2 皇冠分解	19
3.2.1 点覆盖与皇冠分解	21
3.2.2 P_2 -Packing 与皇冠分解	23
3.3 极值归纳技术	31
3.3.1 极值归纳技术的基本原理	31
3.3.2 边不相交三角形 Packing	32
3.3.3 最多内部节点生成树	34
3.4 随机方法	36
3.5 基于低度点的核心化方法	38
3.5.1 基于低度点核心化方法的基本思想	38
3.5.2 连通点覆盖问题的核	40
3.5.3 边支配集	44
3.6 核下界技术	45
3.6.1 对偶性方法	46
3.6.2 基于复杂性理论假设的方法	47
3.6.3 基于参数化规约	49

3.7 本章小结	50
第4章 分支搜索法	51
4.1 常规的分支搜索法	52
4.2 基于隐含参数的分支搜索法	55
4.3 核心化-分支交替搜索法	58
4.4 基于组合的分支搜索法	60
4.5 本章小结	64
第5章 迭代压缩和局部贪婪	65
5.1 迭代压缩	65
5.1.1 提出背景与技术要点	66
5.1.2 典型应用与效率探讨	70
5.2 局部贪婪	73
5.2.1 基于极大解和目标解关系的局部贪婪	73
5.2.2 基于 k 大小目标解求解 $k+1$ 大小目标解的局部贪婪	77
5.3 递归压缩和局部贪婪的运用	78
5.4 本章小结	80
第6章 随机参数算法设计技术	81
6.1 随机方法种类及其应用	82
6.1.1 基于划分的随机方法	83
6.1.2 基于分块的随机方法	88
6.2 确定化方法	89
6.2.1 (n, k) -Universal Set	89
6.2.2 3-Set Packing 随机算法的确定化	90
6.3 本章小结	93
第7章 彩色编码	94
7.1 基本概念	95
7.2 构造方法	95
7.2.1 随机化构造方法	96
7.2.2 确定化构造方法	98
7.3 本章小结	114
第8章 平面图参数算法设计技术	116
8.1 平面图上的核心化技术	116
8.1.1 平面图的基本概念	116
8.1.2 区域分解技术的基本原理	118

8.2 基于平面图的参数算法设计	122
8.2.1 平面支配集问题的亚指数算法	123
8.2.2 层状分割性质与平面图问题参数算法	125
8.3 本章小结	127
第9章 固定参数枚举	129
9.1 固定参数枚举理论	129
9.2 基于分支搜索的枚举	130
9.3 基于彩色编码的枚举	136
9.4 基于递归压缩的枚举	141
9.4.1 FVS的固定参数枚举子过程	143
9.4.2 FVS的固定参数枚举算法	153
9.5 本章小结	155
第10章 参数算法与近似算法	156
10.1 广义参数化近似算法	157
10.1.1 常数近似率广义参数化近似算法	157
10.1.2 固定参数可解时间近似方案	160
10.1.3 以近似性能 $1/\epsilon$ 作为参数的近似算法	162
10.2 标准参数化近似算法	162
10.2.1 常数近似率标准参数化近似算法	163
10.2.2 函数近似率标准参数化近似算法	165
10.2.3 固定参数可解问题的近似算法	168
10.3 3-D Matching 计数问题的一种参数化随机近似算法	170
10.3.1 算法的基本思想	170
10.3.2 算法的主要步骤	171
10.4 不存在参数化近似算法的问题	174
10.5 本章小结	175
第11章 树分解及其应用	176
11.1 树分解基本理论	176
11.2 基于树分解的参数算法设计	178
11.3 其他宽度参数	180
11.4 树分解的应用	181
11.5 本章小结	187
第12章 参数计算的实际应用	188
12.1 单体型计算问题	188
12.2 生物多叉系统发生树最大一致森林问题	195

12.3 无线网络中延时受限的最小能量组播路由的参数算法研究·····	213
12.4 本章小结·····	226
参考文献·····	227
附录·····	234

第 1 章 导 引

城市 A 由三个城区 B、C、D 构成,由于城区 B 离市中心较远,其犯罪率比较高,经常有偷盗、抢劫等案件发生。为了创建文明城市,市政府决定对城区 B 加强治安管理,采取的措施是在某些交叉路口处安装监控设备实现对整个城区的道路实时监管。为了节约成本,市政府要求安装最少的监控设备实现对整个城区道路的监管。市政府决定对该项目进行公开招标,要求投标公司给出节约成本的具体解决方案。在投标这一天,三家公司对该项目抱有极大的兴趣,前来投标。市政府投标主持人说:“大家好,今天对城区 B 的交叉路口监控设备安装项目进行投标。我们市的城区 B 一共有 n 个交叉路口,请问在 k 个交叉路口安装监控设备能否实现整个城区的监管? 请各公司的相关负责人讲述自己的解决方案,哪个方案能够节约成本,并且工期合理,将获得城区 B 的监控设备安装项目。”公司 1 的负责人说:“既然城区 B 一共有 n 个交叉路口,我们的策略是对任意 k 个交叉路口都试一下,确定在这 k 个交叉路口安装监控设备能否实现整个城区的监管。这样,一定能够找到成本最少的方案。”公司 2 的负责人说:“公司 1 提出的方案我们也考虑过,但是从 n 个交叉路口中把所有的 k 个交叉路口都试一遍,工期太长。我们的方案是,对于任意一条没有被监控的道路,在道路的两个路口处都装上监控设备。基于该策略,最后所有的道路都可以被监管。成本可能会增加一倍,但优点是工期很短。”公司 3 的负责人说:“我们观察到,要实现对城区 B 的监管,在城区 B 中安装监控设备的交叉路口数不超过 60。我们的方案是,对于任意一条没有被监控的道路 L ,假设道路 L 的两个路口为 a 、 b ,要么在路口 a 安装一个监控设备,要么在路口 b 安装一个监控设备。找到所有的安装在 k 个交叉路口的可能性,确定在 k 个交叉路口安装监控设备能否实现整个城区的监管。如果存在 k 个交叉路口,使得在这些路口安装监控设备能实现整个城区的监管,那么我们的方案能够找到这 k 个交叉路口且成本最少。另外,我们方案的工期虽然比公司 2 的方案差,但比公司 1 的方案好许多倍。”市政府相关负责人听取了三个公司的汇报,考虑了工程的成本和工期,最终决定让公司 3 负责城区 B 的监控设备安装。

在上述事例中,如果将每条道路看成一条边,道路两端的交叉路口看成顶点,则城区 B 的交通道路可以抽象成一个图 $G=(V,E)$ 。因此,在 k 个交叉路口安装监控设备能否实现对整个城区监管的问题就变成在图 G 中找 k 个点的集合 V' ,使得对于图 G 中的任意一条边 e ,至少 e 的一个端点被包含在 V' 中,即参数化点覆盖问题。

定义 1.1 参数化点覆盖问题。

输入: 图 $G=(V,E)$ 和正整数 k 。

输出: 一个大小不超过 k 的点覆盖, 或返回 G 中不存在大小不超过 k 的点覆盖。

点覆盖问题是一个 NP 难问题, 三个公司提出求解该问题的三种方案。公司 1 的方案事实上是对点覆盖问题的精确求解, 可直接得到时间复杂度为 $O(n^k)$ 的算法, n 为给定图的大小。很显然, $O(n^k)$ 的算法是不能被人们接受的。公司 2 提出的方案是求解点覆盖问题的近似算法。假设点覆盖问题的解为 C 。该近似算法可得到一个大小为 $2|C|$ 的点覆盖, 很好地对点覆盖问题进行了近似求解。然而, 该近似解并不能满足实际应用的精确性和成本的要求。公司 3 通过城区 B 的实际应用背景进行研究, 发现城区 B 中安装监控设备的交叉路口数不超过 60。基于这一参数特性, 提出时间复杂度为 $O(2^k + kn)$ 的参数算法, 其中 n 是给定图的大小。不难看出, 上述算法实现了对点覆盖问题的精确最优求解。基于对参数 k 的观察, 该算法在人们可接受的时间内求解理论上根本不可能被求解的问题。为了进一步比较一般精确算法与参数算法, 下面给出 n, k 取不同值时, n^k 与 $2^k n$ 两类算法复杂性之间的明显区别(用值 $n^{k-1}/2^k$ 表示), 如表 1.1 所示。

表 1.1 n, k 为不同值时 $n^{k-1}/2^k$ 的具体数值

参数	$n=50$	$n=100$	$n=150$
$k=2$	625	2500	5625
$k=3$	15 625	125 000	421 875
$k=5$	390 625	6 250 000	31 640 625
$k=10$	1.9×10^{12}	9.8×10^{14}	3.7×10^{16}
$k=20$	1.8×10^{26}	9.5×10^{31}	2.1×10^{35}

由此可以看出, 对于传统的难解问题, 在理论上是根本不可能有效求解的。在实际应用中, 可以根据实际应用背景引入某些参数对问题进行重新刻画, 而所取参数在实际应用中只在一个小的范围内变化, 通过充分利用小参数这一特殊性质, 可快速解决相应的难解问题。这正是本书介绍的参数计算理论的基本思想。

下面给出其他例子分析参数计算方法的优越性和可用性。

例 1.1 二分图受约束点覆盖问题

在超大规模集成电路领域, 随着 VLSI 技术的高度和阵列密度的增加, 制造过程中缺陷出现的可能性也随之增加。另外, 在产品处理阶段, 由于集成密度的增大, 处理器损坏的可能性也随之增高。为了应对上述问题, 人们引入了容错技术。VLSI 中容错技术对集成电路的发展有十分重要的意义和应用价值。现在较为常用的容错方法是在存储器内部集成备用行和备用列以替换阵列中有缺陷的行

和列。假设用 ku 表示备用行的数目,用 kl 表示备用列的数目。上述容错问题可抽象为用 ku 个备用行和 kl 个备用列对一个大小为 $M \times N$ 的有缺陷阵列进行修复。

人们将上述问题转化为二分图受约束点覆盖问题。

定义 1.2 二分图受约束点覆盖问题。

输入:二分图 $G=(U,L,E)$ 和正整数 ku,kl 。

输出:两个子集 $C_1 \subset U, C_2 \subset L$, 满足 $|C_1| \leq ku, |C_2| \leq kl$, 且使得 E 中每条边至少有一个端点在 $C_1 \cup C_2$ 中, 或返回不存在。

二分图受约束点覆盖问题已被证明是一个难解问题。为了解决该问题,人们提出各式各样的启发式算法和近似算法,但都不能满足实际应用的精确要求。

在超大规模集成电路的实际容错应用中,备用行和备用列的个数之和($ku+kl$)通常小于 40,该值远小于阵列中的行数和列数。于是,人们引入参数 ku 和 kl 对问题重新进行刻画,设计时间复杂度为 $O(1.182^{ku+kl})$ 的参数算法,实现了对问题的有效求解。上述算法较好地解决了超大规模集成电路中的容错问题。

例 1.2 程序类型检测问题

程序类型检测问题是指给定一段 ML 程序 P , 问 P 中类型声明是否一致。该问题在程序设计等领域中有着重要的应用。从难解性的角度来说,该问题比大多数的难解问题(点覆盖、二分图受约束点覆盖等等)难。然而,人们观察到在计算机科学中,程序的嵌套深度一般不超过 10。于是,人们以程序嵌套的深度为参数对问题重新进行刻画,使得对于长度为 n , 嵌套深度为 k 的程序,程序类型检测问题可在 $O(2^k)$ 时间内求解。可见,通过引入恰当的参数,可以在实际应用中解决一些理论上比难解问题更难的问题。

例 1.3 单体型组装问题

在生物信息计算中,单核苷酸多态性是一个物种中不同个体表型的主要遗传来源。识别 SNP 对基因的精确定位,了解基因功能对遗传病等疾病的诊断和药物研究有重要作用。

一个 SNP 位点指的是在一个物种基因组 DNA 序列中不同个体可能出现不同碱基的位置。在一条染色体 SNP 位点上的碱基序列叫做单体型。人类等二倍体生物的染色体是成对存在的,都有一对单体型。

单体型组装问题是指给定某个个体一组已测序的 DNA 片段数据,要求去掉最小数量的数据,以便发现一对单体型与剩下的数据兼容。

人们先前在此问题的研究中,已经意识到这个问题的难解性,并提出解决这一问题的启发式算法,但效果不好。后来,人们对真实的单体型已有数据进行了研究,发现用目前最流行的 DNA 直接测序法,一个长度 1000bp 的片段上的 SNP 位点数通常在 10 个以内。另外,片段数据的覆盖度通常在 5 左右。针对这些真实生

物数据的特点,以 SNP 位点数和片段数据的覆盖度两个量作为参数对单体型组装问题进行建模得出的一系列比原有算法快得多又准确的算法。

本书将从参数计算理论、参数算法设计和分析技术、参数算法实际应用等方面系统介绍参数计算的理论框架和技术框架。具体的,本书将对固定参数可解、固定参数不可解、核心化、迭代压缩、局部贪婪、彩色编码、随机参数化方法、分支界定、图分解、平面图参数算法、固定参数枚举、参数计算的应用等方面进行介绍,通过具体的例子阐述参数计算理论、方法的精髓和具体应用,旨在使读者能够全面系统的认识参数计算理论和技术,掌握相关技术的应用,从而有效求解各领域中的难解问题。

第 2 章 参数计算简介

在计算机科学领域, NP 难解理论为人们奠定了计算难解的界限, 其将所有可计算问题, 划分为“实际可计算问题”(P 问题或易解问题) 和“实际不可计算问题”(NP 难问题或难解问题)。该理论在过去的近四十年中主导着人们在算法设计与分析中的研究。一旦一个计算问题被证明是 NP 难解的, 这一问题就被认为是无法用现代计算机求解的问题, 从而避免了人们为解决此类问题而付出大量而又没有希望的努力。参数计算理论对 NP 难解问题进行了重新划分, 将难解问题分成固定参数可解问题和固定参数不可解问题。可以说, 参数计算理论的提出重新指导了人们求解难解问题的思路和方法。本章将简要介绍 NP 完全理论的相关定义。然后, 对参数计算理论中的固定参数可解、固定参数不可解框架、固定参数枚举等方面进行简要介绍。最后, 对参数计算领域中常用的参数化方法进行简介。

2.1 NP 完全理论

实际工程应用中存在诸多计算优化问题, 有些问题很容易求解, 但有些问题却很难。例如, 张三想驾车从 A 地到 B 地, 从 A 到 B 有很多可选路线且中间经过很多城市, 假设驾车的花费只跟路线的长度有关, 为了节省费用, 张三需要选择一条最短的路线。该问题就是著名的最短路径问题。张三很幸运, 他能够在多项式时间内找到一条从 A 到 B 的最短线路。又如, 李四想去 n 个不同的城市旅游, 因为旅游费用有限, 所以他想找到一个旅游路线既经过这 n 个城市, 又费用最少。上述问题就是著名的旅行商问题。然而, 李四不够幸运, 他很难找到一个最省钱的策略。通过以上两个例子可以看出, 最短路径问题和旅行商问题具有不同的难解性, 前者可以在多项式时间内被求解, 后者却无法利用现有的计算资源进行求解。NP 完全理论^[1]很好地对上述问题进行了分类: 最短路径问题属于 P 问题, 即多项式时间内可解问题, 旅行商问题属于 NP 难解问题。

下面给出 P、NP、NP 完全、NP 难的定义。

首先要说明的是, 在 NP 完全理论框架内, 考虑的问题形式都是判定性问题, 即回答“是”或“否”的问题。对于上面给出的最短路径问题和旅行商问题, 都是最优化问题, 即寻找满足某一特性的最大或最小值。下面以旅行商问题为例具体说明最优化问题和判定性问题的差别。最优化形式的旅行商问题是给定完全图 $G=(V, E)$, 其中每条边赋有一定的权值, 寻找一条权值最小且包括图中所有点的

圈。对于旅行商问题,其判定性形式是给定完全图 $G=(V,E)$,其中每条边赋有一定的权值,并给定参数 K ,问 G 中是否存在一条权值小于等于 K 且包括图中所有点的圈。可以看出,把最优化问题变成判定性,只需要引入一个与问题目标解相关的参数,然后问给定实例是否存在与引入参数相关的解。下面给出判定性问题的具体定义。

定义 2.1(判定性问题) 对于给定问题的每个输入实例,都存在某算法能给出 Yes(是)或 No(否)的判定性回答,则称该问题为判定性问题。如果某算法能对一个输入实例回答为 Yes,那么这个输入实例就叫做 Yes 实例,反之叫做 No 实例。

如前所述,最短路径问题是多项式时间内可解的,其判定性问题是 P 问题。下面给出 P 问题的严格定义。

定义 2.2(P) 如果给定问题 Q 能在多项式时间被求解,则称判定性问题 Q 是 P 问题。

可以说,P 问题就是能在多项式时间内被确定性计算模型,如图灵机解决的判定性问题,如单源最短路径、最小生成树和最大流问题等。

相对于 P 问题的定义,NP 问题非正式定义为,目前来说不能在多项式时间内被确定性计算模型解决,但能在多项式时间内被迅速验证的问题,或者说这类问题能在多项式时间内被非确定性计算模型解决。具体定义如下^[1]。

定义 2.3(NP) 如果一个判定性问题 Q 能被一个多项式时间算法 A 验证,且满足如下性质:

① 对于问题 Q 的 Yes 实例 x ,有一个长度能被多项式 $p(|x|)$ 限定的 2 进制字符串 y ,对于输入元组 (x,y) ,算法 A 回答 Yes。

② 对于问题 Q 的 No 实例 x ,任何长度能被多项式 $p(|x|)$ 限定的 2 进制字符串 y ,对于输入元组 (x,y) ,算法 A 回答 No。

则称该问题属于 NP。

在 NP 完全理论框架中,多项式规约在 NP 难的定义和问题 NP 难解性的证明等方面有重要的作用。下面给出具体的定义。

定义 2.4(多项式多对一规约) 给定两个判定性问题 Q_1 和 Q_2 ,如果存在一个在多项式时间内可计算的函数 r ,对于 Q_1 的任何实例 x , x 是 Q_1 的 Yes 实例当且仅当 $r(x)$ 是 Q_2 的 Yes 实例,则称问题 Q_1 可多项式多对一规约到问题 Q_2 ,记为 $Q_1 \leq_m^p Q_2$ 。

基于多项式规约,可得 NP 难和 NP 完全的定义。

定义 2.5(NP 难) 给定问题 Q ,如果 NP 中的任何问题都能在多项式时间内多对一规约到问题 Q ,则称判定性问题 Q 是 NP 难的。

定义 2.6(NP 完全) 如果一个判定性问题 Q 是 NP 难的,同时又在 NP 中,