Combinatorial Algorithms for Batch Scheduling and Network Problems

批调度与网络问题的组合算法

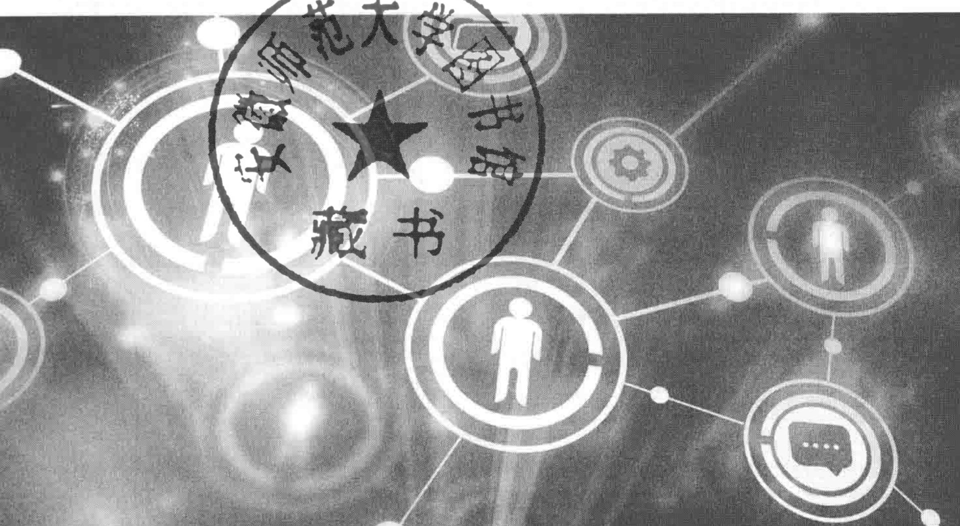
李曙光　于立萍　宋英杰　张　斌◎著

中国工信出版集团　人民邮电出版社
POSTS & TELECOM PRESS

Combinatorial Algorithms for Batch Scheduling and Network Problems

批调度与网络问题的组合算法

李曙光　于立萍　宋英杰　张　斌◎著



人民邮电出版社
北京

图书在版编目（CIP）数据

批调度与网络问题的组合算法 / 李曙光等著. -- 北京 : 人民邮电出版社, 2017.7
ISBN 978-7-115-45595-6

Ⅰ. ①批… Ⅱ. ①李… Ⅲ. ①调度程序 Ⅳ. ①TP315

中国版本图书馆CIP数据核字(2017)第098871号

内 容 提 要

本书以作者在算法设计领域的研究成果为基础，给出了求解批调度问题的一系列组合算法，以及求解网络优化问题的若干组合算法。主要研究了极小化加权完工时间和、最大延迟和最大完工时间 3 种调度目标函数，以及网络中的呼叫接纳、利润极大化和 t 区间的 k 染色问题等。

本书可作为从事调度理论、组合最优化、算法设计与应用科技人员的参考书。

◆ 著　　　李曙光　于立萍　宋英杰　张　斌
　责任编辑　邢建春
　责任印制　彭志环

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
　邮编　100164　　电子邮件　315@ptpress.com.cn
　网址　http://www.ptpress.com.cn
　三河市海波印务有限公司印刷

◆ 开本：880×1230　1/32
　印张：3.5　　　　　　2017 年 7 月第 1 版
　字数：95 千字　　　　2017 年 7 月河北第 1 次印刷

定价：39.00 元

读者服务热线：(010) 81055488　印装质量热线：(010) 81055316
反盗版热线：(010) 81055315

前 言

优化是根据现状采取行动以获得最好（或最优）的结果。组合最优化研究的是可行解数目有限的离散优化问题。在复杂性理论的框架下，希望能在输入规模的多项式时间之内找到最优解。运行在多项式时间之内的算法称为有效算法。输出最优解的有效算法称为精确（或最优）算法。

当多项式时间精确算法很可能不存在时(所谓的NP难解问题)，转而设计有效的近似算法以得到次优解。对于极小化问题，算法的近似比（性能比）定义为算法给出的解的目标值与最优值之间的最坏情形比。对于极大化问题，算法的近似比（性能比）定义为最优值与算法给出的解的目标值之间的最坏情形比。近似比为ρ的算法称为ρ-近似算法。通常用近似比来衡量算法的性能：近似比越小，算法越好。

一族算法$\{A_\epsilon\}$称为一个多项式时间近似方案(PTAS，Polynomial Time Approximation Scheme)，如果对于每一个给定的正数ϵ，算法A_ϵ是一个运行于输入规模的多项式时间之内的$(1+\epsilon)$-近似算法。注意到ϵ是给定的，因此算法的运行时间可以以任意方式依赖于$1/\epsilon$。如果运行时间也是$1/\epsilon$的多项式，就得到了全多项式时间近似方案（FPTAS，Fully Polynomial Time Approximation Scheme）。对于NP

难解问题和所谓的强 NP 难解问题来说，全多项式时间近似方案和多项式时间近似方案分别是目前所能得到的最好结果。

本书研究批调度和网络中的若干确定性优化问题，给出了有效的组合算法。确定性是指问题实例的所有参数均事先已知。研究的问题大多是 NP 难解的，所给出的算法大多是多项式时间的近似方案，其余的是精确算法或常数近似比算法。下面简略地描述所研究的问题，并给出主要结果和创新点。

第 1 章介绍了研究背景；第 2~5 章是第一部分，给出了若干批调度问题的组合算法；第 6~8 章是第二部分，给出了若干网络问题的组合算法。

一个典型的调度问题包含 n 个工件和 m 台机器。在满足一定的约束条件下，要将工件安排在机器上进行加工，目标是找到一个最优的安排。最优性是由依赖于问题的目标函数所定义的。本书研究分批调度问题。每台机器可以同时加工若干个工件，这些工件构成一个批次，这样的机器称为批机器。将工件分批加工是为了提高效率，分批加工比逐个加工更快或成本更低。

批调度问题有多种模式。本书研究如下的煅烧模式。给定 n 个工件和 m 台并行的同型批机器。每个工件有一个加工时间，描述了在任意一台机器上加工该工件所需要的最短时间。每个工件还有一个释放时间，在该时间之前不能开始加工这个工件。一个批次的加工时间是该批次所包含的所有工件的加工时间的最大者。一个批次的完工时间等于它的开工时间加上它的加工时间。同一批次中的所有工件有相同的开工时间，也有相同的完工时间，即该批次的完工时间。一个批次从开始加工到完工，不允许向内添加工件或向外移

除工件。每个批次的加工都是连续进行的，即从开工到完工，没有其他批次可以在同一台机器上加工。煅烧模式起源于大规模集成电路生产过程中的煅烧操作调度问题。

煅烧模式分为两类。有界模式：每个批次最多可容纳的工件数目（批容量）B 小于工件的总数目 n，B 表示任一台机器能同时加工工件的最大数目。无界模式：每个批次可容纳的工件数目没有限制，即 $B \geqslant n$。无界模式的一个例子：在干燥炉中硬化一些化合物，干燥炉足够大，因此批容量没有限制。注意有界模式中 $B=1$ 的情形，就是经典的调度问题（每台机器在任一时间至多加工一个工件）。因此，有界模式的批调度问题，难度不低于经典的调度问题。

集中研究工件释放时间不相同的调度问题。这比所有工件同时到达的问题要难得多。研究 3 种调度目标：极小化加权完工时间和、极小化最大延迟、极小化最大完工时间。记工件 j 在一个调度中的完工时间为 C_j。

第 2 章和第 3 章分别研究极小化加权完工时间和的有界批机器和无界批机器并行调度的问题。每个工件 j 有正权 w_j，加权完工时间和，顾名思义，是 $\sum_j w_j C_j$。这两个问题都是 NP 难解的，并且前者是强 NP 难解的。在批调度的各种目标函数中，极小化加权完工时间和是最难解决的问题之一。对于这两个问题，这两章均分别给出多项式时间近似方案。有一点奇怪的是，在所得到的算法中，那些处理批调度问题的技巧用得并不多，反而更多地依赖于研究经典调度问题（$B=1$）所发展起来的技巧。

第 4 章主要研究极小化最大延迟的有界批机器并行调度问题。每个工件 j 最好能在它的交货期 d_j 之前完工。最大延迟定义为

$\max_j\{C_j - d_j\}$。由于最优值可能为负值，在这一章中，应用了最大延迟问题的等价形式——送货时间表述法。该方法为每个工件 j 设了一个送货时间 $q_j = \max_l d_l - d_j$，这样目标就变成极小化 $\max_j\{C_j + q_j\}$。这个问题是强 NP 难解的。该章给出了第一个多项式时间近似方案。所用技巧经过简单修改之后，就得到（NP 难解的）极小化最大延迟的并行机无界批调度问题的第一个多项式时间近似方案。当所有的 d_j 为 0 时，最大延迟就是最大完工时间 $\max_j C_j$。因此也得到了求解极小化最大完工时间的有界批机器和无界批机器并行调度问题的多项式时间近似方案。

第 5 章研究了工件具有不同尺寸的单机批调度问题，目标函数是极小化最大完工时间。每个工件 j 有一个尺寸 $s_j \in (0,1]$。批机器可以将若干个工件作为一批同时进行加工，只要这些工件的尺寸之和不超过 1。工件尺寸不同的调度问题是批调度问题的一般形式。显然，此时不必考虑无界模式。最大完工时间，正如上面所定义的，是调度中所有工件完工时间的最大者。对于这个问题，给出了 $(2+\epsilon)$-近似算法，ϵ 是任意小的正数。算法的运行时间是 $O(n\log n) + f(1/\epsilon)$，隐藏在 $O(n\log n)$ 中的常系数很小并且与 ϵ 无关。当所有工件同时到达并且加工时间相同时，这一问题就是经典的一维装箱问题。一维装箱问题是强 NP 难解的，近似比小于 3/2 的近似算法很可能是不存在的。

除了批调度问题，还研究通信网络中出现的优化问题。通信网络在社会经济生活中日显其重要性，因此关于网络设计与有效运营的优化问题受到了业界的普遍关注。除去大量的实际应用之外，网络问题也有很多有趣的方法论特性。作为一种流行的通信网络，环

形网引起了很多人的兴趣，人们开始集中研究环形网。

在第 6 章，考虑环形网中的呼叫接纳控制问题。给定一个环（无向或有向）和一组通信请求。每个请求用一对顶点（无序或有序）表示，并且有一个利润。要在无向环中实现一个呼叫，需在环中指定该请求所对应的两个顶点之间的一条路。要在有向环中实现一个呼叫，需在环中指定该请求的源节点到目的节点的一条路。环（无向或有向）网络呼叫接纳控制问题的目标是确定最大利润的呼叫子集，在环中实现该子集中的每一个呼叫，使经过每一边（无向或有向）的呼叫的数目不超过该边的容量。对于无向和有向环形网呼叫接纳控制问题，均首次给出了多项式时间近似方案。

在第 7 章，研究多纤 WDM 网络中的利润极大化问题。要在多纤 WDM 网络中实现一组传输请求，为其中每一个请求安排一条路并分配一个波长，使任一链路上任一波长的使用次数不超过该链路上的光纤数。给定一个波长数目有限的多纤 WDM 网络和一组传输请求，利润极大化问题的目标是要确定利润最大的并且可以在网络中实现的那一部分传输请求。第 6 章所研究的呼叫接纳控制问题，是这一问题只有一个波长时的特例。对于链网，给出了多项式时间精确算法。环形网中的这一问题是 NP 难解的，即使所有传输请求的利润均为 1 时也是如此。给出了两个算法，近似比分别为 2 和 $1.582+\epsilon$，ϵ 是任意小的正数。对于环上各边光纤数目相同的均匀模式，给出了 1.582-近似算法。这些结果也适用于有向链网与环形网。注意在将多纤环形网中的 2-近似算法推广到有向多纤环形网时，要求某一链路上的两条有向边分别为顺时针和逆时针方向上容量最小的有向边。

最后，在第 8 章引入了圈上t-区间的k-染色问题。这一问题推广了 WDM 环形网中几个熟知的优化问题。圈代表环形网，颜色代表波长，t-区间代表客户请求。一个t-区间，由圈上至多t（$t\geqslant 1$）个区间构成，每个区间的两个端点都是圈上的顶点。要实现一个客户请求，需选择它所对应的t-区间中的一个区间并为其安排一种颜色。在实现任意两个请求时，所选定的两个区间如果在圈上有公共边，则不能得到同一种颜色，否则会引起波长冲突。给定一个圈、k种颜色和若干个请求（t-区间），问题的目标是实现最大数目的请求。我们给出了这一问题的一个 3.042-近似算法，顺便也得到了链网中t-区间的k-染色问题的一个 2.542-近似算法。

本书的写作得到了山东大学李国君教授的帮助与支持，出版得到了山东省高校智能信息处理重点实验室（山东工商学院）的资助，在此表示衷心的感谢！

本书可作为从事调度理论、组合最优化、算法设计与应用科技人员的参考书。

由于学术水平有限，书中难免有错误和疏漏之处，希望能得到各位学者的指正与帮助。有任何的问题需要讨论请联系 sgliytu@hotmail.com。

李曙光

2017 年 3 月

目 录

第1章 绪 论

1.1 背景知识

调度问题起源于有限资源的合理分配。目标是找到一个最优的分配方案，最优性是由依赖于问题的目标所定义的。在这样一个普遍的定义之下，自然地，调度是最优化的一个重要子领域。在经典调度理论中[1]，每台机器同一时间至多加工一个工件。近几十年来，批调度问题得到了相当多的研究。一台批机器（或称批加工机器，简称 BPM，Batch Processing Machine）能将若干个工件作为同一批次同时进行加工[1]。要了解批调度的研究进展，可参阅文献[2~4]。

Webster 等[5]区分了三类分批调度：一个批次的加工时间等于该批次中所有工件加工时间的最大者，称为煅烧模式（参阅文献[6]）；一个批次的加工时间等于该批次中所有工件的加工时间之和（参阅文献[7]）；一个批次的加工时间是一个常数，与该批次中所包含的工件无关（参阅文献[8]）。

煅烧模式起源于大规模集成电路生产过程中的煅烧操作调度

问题。集成电路生产的最后一个阶段是煅烧操作，即将芯片放在板子上然后放入烤箱烘烤。烘烤过程中出现问题的芯片就被淘汰。每个芯片有一个预定的煅烧时间，经受住这段时间烘烤的芯片被视作合格产品。芯片的煅烧时间与其类型和（或）客户的要求有关。由于芯片在烤箱中的停留时间可以大于其煅烧时间，因此可将不同的产品作为一批同时放入烤箱进行烘烤。这样，一个批次的加工时间就是该批次中所有产品煅烧时间的最大者。将烤箱视为批机器。煅烧操作调度问题，就是将所有的n个工件进行分批，然后在机器上安排加工，使某个目标函数最优。

研究了两类煅烧模式。（1）有界模式：每个批次最多可容纳的工件数目（即批容量（B））小于工件的总数目n。（2）无界模式：每个批次可容纳的工件数目没有限制，即$B \geqslant n$。无界模式的一个例子：在干燥炉中硬化一些化合物，干燥炉足够大，因此批容量没有限制。若无特别说明，本书提到批机器时，指的就是有界批机器。

集中研究工件释放时间不相同的调度问题，这比所有工件同时到达的问题要难得多。研究 3 种调度目标：极小化加权完工时间和、极小化最大延迟、极小化最大完工时间。要了解调度目标函数的定义，请参阅文献[9]。

除了批调度问题，本书还研究通信网络中出现的优化问题。通信网络在我们的社会经济生活中日显其重要性，因此，关于网络设计与有效运营的优化问题受到了学术界的普遍关注。除去大量的实际应用之外，网络问题也有很多有趣的方法论特性。

网络模型由两部分构成：边（有时称为弧）和顶点（或节点）。边是连接线，顶点是边的连接点。图 $G=(V,E)$ 是通过 E 中的边将 V 中的顶点互连所得到的结构，V 和 E 分别代表顶点集和边集。顶点代表处理器，边代表处理器之间的通信链路。有向图是指其中的弧有指定方向的图。弧没有指定方向的图称为无向图。网络是一个边赋权图（边上所赋的权有时称为容量），通常可分为有向网络和无向网络。

G 中一列连续的边称为路(链)。将路的两个端点粘合在一起，就得到了一个圈（环）。若对任意一对节点 $u,v\in V$，G 中有一条从 u 到 v 的路，则 G 是连通的。若 G 是连通的并且不含圈，则称为树。链、环、树是实际应用中基本的网络拓扑结构，很多其他的网络是这些网络结构的组合或者派生。

波分复用（WDM，Walength Division Multiplexing）有效利用了光网络所提供的巨大带宽，是目前广泛应用的技术。波分复用技术的基本原理是把一根光纤的带宽分成多个信道，每个信道分配一个不同的波长，以使不同信号能在同一光纤的不同信道上同时传输[10,11]。要实现一个传输请求，必须为该请求建立连接，即在网络中选择一条从源节点到目的节点的路并分配一个波长，然后光信号便沿着该路传输而无需改变波长，从而避免了光—电—光转换所带来的消耗和延迟[10]，这样的网络称为全光网络。为了避免波长冲突，一根光纤上不允许多个传输请求同时使用同一波长。在多纤网络中，经过同一链路的多个传输请求如果由不同的光纤携载，则允许同时使用同一波长，而不会出现波长冲突。这样的

网络应用了光交叉连接技术，能够将任一输入链路某根光纤上的信号输出到指定输出链路上另一根光纤上而无需改变波长[12]。

在实际应用中，带宽总是有限的，当前技术仅能提供几百个波长。因此，未必所有请求都能被实现。这样一来，极大化被实现的请求数目就显得很重要了。从网络供应商的角度来说，更重要的是利润极大化[13,14]。

受到这一类应用的启发，研究如下的利润极大化问题。设给定一个具有有限数目波长的多纤 WDM 网络和一组点—点通信请求，每个请求联系一个利润。对于一组传输请求，如果能在网络中为其中每一个请求选择一条路并分配一个波长，使任一链路上任一波长的使用次数不超过该链路上的光纤数，则称该组传输请求可以在网络中实现。利润极大化问题的目标是要确定利润最大的并且可以在网络中实现的那一部分传输请求。本书还研究了两个相关问题。

本书集中研究环形网。在环形网中，每个节点扮演着同样的角色。这种节点的对称性简化了网络算法（如路由与波长分配）的设计。另外，环是 2-连通的，环中任意一对节点之间存在两条不同的路。因此，一个节点或者一条链路出了问题，环仍然保持连通。因为有这些优点，环在通信网络中得到了广泛应用，也引起了很多人的研究兴趣[15~23]。

所研究的大多数问题是 NP 难解的，找不到多项式时间的精确算法，除非 P=NP[24]。因此，把注意力放在寻找多项式时间近似的算法上。考虑到部分问题的实际应用背景，研究的重点是得

到组合算法。组合算法有两个好处：一是简单易于实现；二是经常能提供一些问题的结构特点，以便应用到特殊情形上。

1.2 算法复杂性的若干基础概念

下面先介绍来自复杂性理论的一些基础概念。这些概念是相当技术性的，因此这里只给出要点。

定义 1.1 一个 NP 优化问题 Π 是一个极小化或极大化问题。它的每一个真实的实例 I 伴随着一个非空可行解集，每个可行解有一个目标函数值，该目标函数值是一个非负实数。存在多项式时间算法能够确定真实性、可行性和目标函数值。

定义 1.2 一个优化问题是 NP 难解的，如果所有的 NP 优化问题都能多项式时间归约到它。

定义 1.3 一个优化问题是强 NP 难解的，如果即使当输入实例以一元表示时，它仍然是 NP 难解的。

定义 1.4 如果存在常数 $c>0$，使得当 n 足够大时有 $f(n)\leqslant cg(n)$，则记 $f(n)=O(g(n))$。

定义 1.5 如果一个算法的运行时间在输入规模的多项式时间之内，则称之为有效算法。

定义 1.6 如果一个有效算法能够给出问题的一个最优解，则称之为精确（或最优）算法。

定理 1.1[24] 除非 P=NP，任何 NP 难解问题都没有多项式时间精确算法。

根据这个定理，很自然地，对于 NP 难解问题则要寻找好的近