

调度问题及其解空间的特征分析

孙元凯, 刘 民, 吴 澄
(清华大学自动化系, 北京 100084)

摘 要: 目前, 在组合优化领域中, 评判近优算法的性能尚缺乏统一的标准和有效的依据, 而算法的效率与所要解决问题之间的关系密不可分. 本文以 JSP 问题为例, 研究了调度问题本身的结构特征, 分析了调度问题可行解空间的属性, 提出了分割因子的概念. 研究表明, 分割因子影响调度问题可行解空间的规模, 而各工序加工时间的分布则影响解空间的“崎岖”状况; 分割因子和工件加工时间的分布在一定程度上可以反映调度问题的复杂程度. 这对近优算法的设计具有一定的指导意义, 并为建立统一的近优算法效率衡量标准迈出了探索性的一步.

关键词: JSP; 调度; 复杂性; 原空间; 可行解空间; 分割因子

中图分类号: TP20 **文献标识码:** A **文章编号:** 0372 2112 (2001) 08 1042 04

Analysis of Characteristics of Scheduling Problem and Its Solution Space

SUN Yuan kai, LIU Min, WU Cheng

(Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: There are no universal criteria and foundations to evaluate approximate algorithms in combinatorial optimization fields at present, however, the focusing problems make great difference to the efficiency of algorithms. In this paper, we take the Job shop problem as an example, analyze the characteristics of the scheduling problem and its solutions space, and propose a concept: cutting factor. Our research shows that cutting-factor can decide the size of the solution space, and processing time of the operations can influence the roughness of the solution space; cutting-factor and processing time's distribution can indicate the complexity of the scheduling problem. Our study is an exploring step towards establishing universal criteria for approximate algorithms, and can guide the design of algorithms to some extent.

Key words: job shop problem; scheduling; complexity; original space; solution space; cutting factor

1 引言

在组合优化领域中有多种优化算法和近优算法. 优化算法主要是基于分枝定界方法, 但由于大多数组合优化问题属于 NP 问题, 存在指数爆炸的现象, 所有这类算法在大规模问题上的性能表现不理想. 近年来, 近优算法越来越受到人们的关注, 尤其是一些自适应启发式算法如 GA (遗传算法)、TA (禁忌搜寻) 等, 这些算法在某些领域得到了成功的应用^[5-7]. 然而, 在评判这些算法的性能上, 尚缺乏有效的手段和统一的标准. 我们认为, 算法性能的好坏, 除了与算法本身的因素有关外, 还与算法要处理的具体问题有着密切的关系; 同一算法, 其性能在不同的问题上可能表现出很大的差异. 因此, 研究问题本身的复杂性, 对于评估算法性能以及指导算法的设计都具有重要的意义. 本文以 Job Shop 调度问题 (简称 JSP) 为例, 研究了问题本身的结构特征, 分析了调度问题可行解空间的属性, 提出了分割因子的概念. 研究表明, 分割因子影响调度问题可行解空间的规模, 而各工序加工时间的长短则影响局部极值点的分布情况. 分割因子和工件加工时间的分布在

一定程度上可以反映调度问题的复杂程度. 这些工作对于理解调度问题本身的复杂性以及评价调度算法性能等方面都具有一定的参考价值, 并为建立统一的近优算法效率衡量标准迈出了探索性的一步.

本文以 JSP 问题为研究对象的主要原因有两个: 一是 JSP 问题表述简单、求解算法易于实现; 另一个更重要的原因是, 对 JSP 问题, 已经有公认的生成复杂程度不同实例的方法^[8], 从而可以增加算法分析和评价的客观性.

2 JSP 问题的描述

JSP 问题可以简单地描述如下: 有 n 个工件和 m 台机器设备; 每个月都包含 m 个操作 (或称工序), 每个操作都需要单独占用某台设备进行加工; 操作一旦开始, 就不能中途被打断; 每台机器在任一时刻最多只能加工一个工件; 每个工件占用任一台设备的次数不多于一次. 所谓调度, 就是在每台机器上为每个操作分配一段加工时间, 以便寻找一种可行的调度方案, 使得所有工件的完工时间最短. 一个调度方案也称为 JSP 问题的一个解 (本文中的解实际上是指各工序之间的加

工序约束关系; 当各工序之间的加工次序确定后, 就可以推导出各工序的起止时间, 从而得到各工序的排序时间表), 所有可行解构成的集合就称为 JSP 问题的可行解空间(有时也简称为解空间). 为了方便起见, 本文后面的部分将以 $n \times m$ 的形式来表示 JSP 问题的规模, 其中 n 代表工件数量, m 代表机器设备的数量.

3 JSP 问题的原空间和解空间

对某个 $n \times m$ 规模的 JSP 问题, 任取两个解 s 和 t , 定义两解之间的距离为:

$$H(s, t) = \sum_{i=1}^m \sum_{j=1}^{n-1} \sum_{k=j+1}^n \text{XOR}(PREC_{i,s}(j, k), PREC_{i,t}(j, k))$$

其中 $PREC_{i,s}(j, k)$ 表示解 s 中第 i 台机器上工件 j 和工件 k 之间的加工顺序, 若 j 在 k 之前先加工, 则 $PREC_{i,s}(j, k) = 1$, 否则 $PREC_{i,s}(j, k) = 0$; XOR 为逻辑异或运算符. 若 $H(s, t) = 1$, 则称 s 和 t 互为邻居. 对于 $n \times m$ 规模的 JSP 问题而言, 若不考虑解的可行性, 则总共有 $(n!)^m$ 个解, 每个解都有 $m(n-1)$ 个邻居. 若将每个解视为高维空间中的一个点, 在高维空间中对这些解进行排列, 使得每个解都与自己的邻居毗邻, 由此所构成的高维点阵我们称之为 JSP 问题的原空间. 原空间中由可行解组成的集合就是 JSP 问题的解空间. 对所有规模为 $n \times m$ 的 JSP 问题而言, 它们的原空间都是相同的, 但解空间却可能有极大的差别.

4 解空间分析

显然, 影响调度问题复杂程度的因素除问题本身的规模以外, 还有工件的工艺约束路线以及工件加工时间等.

4.1 决定解空间规模的因素

4.1.1 JSP 问题的表述 采用文献[9]中的非连接图模型来描述 JSP 问题, 如图 1 所示. 图中顶点 O_{13} 代表第一个工件的第三个操作, 其他符号类推. 图 1 中, 同一工件的操作, 如 $\{O_{11}, O_{12}, O_{13}\}$, $\{O_{21}, O_{22}, O_{23}\}$, $\{O_{31}, O_{32}, O_{33}\}$ 和 $\{O_{41}, O_{42}, O_{43}\}$, 这些操作之间由有向弧串联, 表示工件的工艺路线. 在同一台机器上加工的所有操作, 如 $\{O_{11}, O_{21}, O_{31}, O_{41}\}$, $\{O_{12}, O_{22}, O_{32}, O_{42}\}$, $\{O_{13}, O_{23}, O_{33}, O_{43}\}$, 这些操作之间以虚线连接. 图 1 中所描述的是 4 个工件在 3 台设备上加工时的非连接图模型. 调度方案就是要确定图中虚线的取向问题. 可行的调度方案可表述为: ①图中所有的虚线均取为单向弧; ②最终的图中不包含回路. 图 2 就是因图 1 中虚线取向单向弧而确定的一种调度方案(图 2 中的虚线弧表示冗余, 可根据其他的单向弧推断出该弧线的取向).

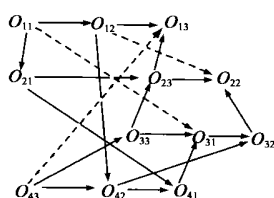
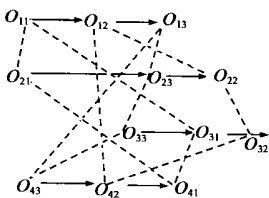


图1 非连接图模型 图2 一种可行的调度方案

4.1.2 解空间规模的粗略估计 对于一个 $n \times m$ 规模的 JSP 问题, 定义如下一个特征数 T :

$$T = \sum_{j=1}^n \sum_{k=j+1}^n \left[\sum_{p=1}^m \sum_{q=p+1}^m \text{XOR}(PREC_j(p, q), PREC_k(p, q)) \right]$$

$PREC_j(p, q)$ 表示工艺约束中工件 j 占用机器 p 和机器 q 的先后次序. 若工件 j 先使用机器 p , 则 $PREC_j(p, q) = 1$, 反之, $PREC_j(p, q) = 0$; XOR 为逻辑异或运算符, 若 $\text{XOR}(PREC_j(p, q), PREC_k(p, q)) = 1$, 则就说工件 j 和 k 在机器 p 和 q 之间存在一个分割数. 将 T 做归一化处理, 得特征值 F :

$$F = T / C_n^2 C_m^2$$

T 称之为分割数, F 称之为分割因子. 分割因子对于问题的解空间有怎样的影响呢?

一个调度方案要可行, 关键是非连接图模型中不得有回路存在. 在图 2 中, 工件 1, 2 在机器 2 和 3 之间存在一个分割数, 若 $\{O_{12}, O_{22}\}$ 和 $\{O_{13}, O_{23}\}$ 之间的弧线方向取反, 则可以看出 $\{O_{12}, O_{22}, O_{13}, O_{23}\}$ 这 4 个工序就构成一个回路, 从而所生成的解就不可行; 若工件 1, 2 在机器 3 和 4 之间不存在分割数, 则这 4 个工序是无论如何也不会形成回路的. 因此, 分割数越大, 任意两个工件之间潜在回路数量就越多. 当任意安排每台机器上工件的加工次序时, 就越容易构成回路, 从而使得可行解数量就越少, 解空间的规模也就越小, 可以粗略估计一下 JSP 问题解空间的规模(即可行解的数量).

对于一个 $n \times m$ 规模的 JSP 问题, 它的分割数为 T , 分割因子为 F , 假设在将虚线取为单向弧时是任意的, 则从平均意义上说, 任意两个工件 n_1 和 n_2 之间的平均分割数为 $T / C_n^2 = FC_m^2$, 平均每台机器所涉及到的分割为 $FC_m^2 / (m/2) = (m-1)F$, 即从平均意义上说, 每台机器与其他 $(m-1)F$ 台机器之间在非连接图模型中有构成回路的可能.

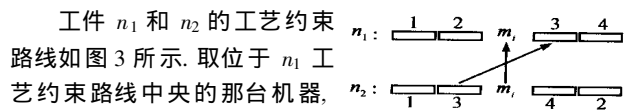


图 3 可行解的构成

平均值, 假设也位于工艺约束路线的中央. 假设在 m_i 机器上先加工 n_2 (图 3 中以向上的箭头来表示), 与 m_i 可构成回路的机器有 $(m-1)F$ 台, 不妨假设其中的 $1/2$ 位于 m_i 之前(标识为集合 2), 另外的 $1/2$ 位于 m_i 之后(标识集合 3); 位于 m_i 之前的其他机器标识为集合 1, 位于 m_i 之后的其他机器标识为集合 4. (标识只表示了各集合的相对排列顺序, 表示集合 1 和集合 2 位于集合 3 和集合 4 之前, 不代表集合 1 位于集合 2 之前, 其他类推. 在 n_2 的工艺约束路线中也是如此). 由于 m_i 上是先加工 n_2 , 为了不形成回路, 标识为 3 的集合中, 每台机器也必须是先加工 n_2 后加工 n_1 ; 集合 1 和集合 3 的机器之间不可能形成回路; 集合 1 和集合 4 的机器之间不可能形成回路; 集合 2 和集合 3 的机器之间不可能形成回路; 集合 3 和集合 4 的机器之间也不可能形成回路. 因此, 由于机器 m_i 上 n_1 和 n_2 加工顺序的确定, 确定了 $(m-1)F/2$ 台其他机器上这两个工件的加工顺序, 剩下的需要考虑集合 1 和集合 2 之间

以及集合2和集合4之间的情况.任取集合2中的一台机器 m_j ,也假设该机器位于中央位置(相对于集合1和集合2而言),采用前面的方法,当确定了 m_j 上 n_1 和 n_2 的加工顺序后,可确定其他 $(m-1)F/4$ 台机器的加工顺序;如此进行下去,就可安排完所有机器上工件 n_1 和 n_2 的加工顺序.若忽略各集合内部形成回路的可能性,可得到不构成回路的概率 P 为:

$$P = \frac{1}{2} \times \left(\frac{1}{2} \right)^{\frac{1}{2}(m-1)F} \times \left(\frac{1}{2} \right)^{\frac{1}{4}(m-1)F} \times \left(\frac{1}{2} \right)^{\frac{1}{8}(m-1)F} \times \dots$$

$$= \frac{1}{2} \times \left(\frac{1}{2} \right)^{(m-1)F}$$

n 个工件两两分组,可以分成 $n/2$ 组,因此,若任意排列每台机器上工件的加工顺序,生成可行解的概率不会高于 $P^{n/2}$,因此,对于 $n \times m$ 规模的JSP问题,可行解空间的规模不超过 $(n!)^m \times P^{n/2}$.由此可见,分割因子越大,可行解空间的规模也就越小,搜索相对较容易.

按照文献[2,8]中的方法生成两组复杂程度不同的实例,规模均为 50×10 ,每组5个实例,分别标识为easy1~easy5和hard1~hard5.这些实例的 T 和 F 值如表1所示.

表1 实例的分割数和分割因子

实例(50×10)	分割数 T	分割因子 F
Easy1	27748	0.503
Easy2	27714	0.503
Easy3	27680	0.502
Easy4	27903	0.506
Easy5	27663	0.502
hard1	12401	0.225
hard2	12231	0.222
hard3	12414	0.225
hard4	12406	0.225
hard5	12343	0.224

从表1中可以看出,难例的分割因子约为0.224,容易实例的分割因子约为0.503.按上面的分析,两者可行解空间规模之比不小于 $0.5^{(m-1)(0.224-0.503)(m/2)} = 2^{62.775} \approx 8 \times 10^{18}$,即难例的可行解空间规模远比容易实例可行解空间的规模大得多.

4.2 局部极值点的分布情况

对于所生成的10个实例,采用随机的方式各自生成1000个解,记为Rand组;然后按最速下降法找到相应的局部极值点,记为Min组. Rand组和Min组各解之间的平均距离如表2所示.

表2 解空间分析

实例(50×10)	Rand组平均距离 (随机生成)	Min组平均距离 (最速下降法生成)
easy1	748	749
easy2	746	747
easy3	738	740
easy4	723	724
easy5	754	756
hard1	1448	1448
hard2	1418	1419
hard3	1401	1401
hard4	1428	1430
hard5	1400	1401

表2中的数据可以说明两个问题:无论对于哪个实例,Rand组和Min组的平均距离几乎一样,这间接反映出解空间中局部极值点的分布是比较均匀的;又因为Rand组的解是随机生成的,hard1~hard5的平均距离均比easy1~easy5的平均距离大,这说明hard1~hard5实例解空间的规模比easy1~easy5实例解空间的规模大,这也验证了在4.1.2节中关于解空间规模问题的推断.

4.3 对于已知实例的测试

为了验证得出的结论,还需要在已知实例上进行测试.选择已经公开发表并且已经知道其最优解的实例;这些实例由不同的作者提出,而且都是同规模的.组1包括实例{orb1, orb3, orb6, orb8},是1986年提出来的(请参阅文献[6]);组2包括{La17, La18, La19, La20},是S. Lawrence在“Resource constrained project scheduling: an experimental investigation of heuristic scheduling technique”一文中提出的.测试中使用的算法是禁忌搜寻算法,以FCFS(先来先服务)规则生成初始解,迭代次数为10000次.对每个实例,分别运行30次,取平均值作为算法的搜索结果,测试结果如表3所示.

表3 已知实例的测试结果

实例	规模	最优解	禁忌搜寻算法的结果	与最优解的差距	分割因子
Orb1	10/10	1059	1172	10.7%	0.270
Orb2	10/10	1005	1135	12.9%	0.174
Orb6	10/10	1010	1135	12.4%	0.270
Orb8	10/10	899	979	8.9%	0.174
La17	10/10	784	811	3.4%	0.503
La18	10/10	848	887	4.6%	0.489
La19	10/10	842	866	2.9%	0.513
La20	10/10	902	939	4.1%	0.481

从表3中可以看出,组1中实例的分割因子平均约为0.22,组2中实例的分割因子平均约为0.5;组1中算法结果的平均差距约为11%,组2中算法结果的平均差距约为4%.从计算结果来看,同样的算法,在不同的实例中所表现出的性能不同:在分割因子小的实例上,搜索结果相对较差;在分割因子大的实例上,搜索结果相对较好.这说明组1实例比组2实例更复杂,求解更困难些.以上测试结果与在4.1节中得出的结论一致:分割因子小的实例相对较复杂,求解较困难.

4.4 各工序加工时间的分布对解空间的影响

以上讨论了工艺约束对JSP可行解空间规模的影响,为了进一步揭示影响JSP复杂度的其它因素,将时间分布对解空间性质的影响做了对比实验.在4.1节中,在生成实验用例时,工件的加工时间在区间[0,100]上随机生成,从Rand组中的解搜索到相应的极值点,算法的平均迭代次数记做step1;由于只改变工件的加工时间并不影响解的规模(即不影响可行解的数量;在本文中,所说的解,只指各工件之间的加工次序约束关系,因为有了加工次序关系,就可以推断出每个工件的加工起止时间),因此,保留Rand组的解,但将所有工件的加工时间都取为相同的值,并搜索其相应的极值点,算法的平均迭代次数记做step2.两组计算结果如表4所示.

表 4 时间分布对解空间的影响

实例 (50×10)	组 1 的平均迭代次数 step1 (各工序加工时间均匀分布)	组 2 的平均迭代次数 step2 (各工序加工时间相同)
Easy1	28.59	1.84
Easy2	19.78	1.51
Easy3	17.31	3.55
Easy4	28.11	1.21
Easy5	21.09	4.13
Hard1	16.84	0.62
Hard2	14.54	0.76
Hard3	15.24	0.63
Hard4	14.84	0.73
Hard5	19.02	1.06

从表 4 中可以看出, step1 明显大于 step2. 若以可行解空间为超平面, 以完工时间表示各解的高度值, 则得到一个高低不平、类似于地表结构的特征图. 在该特征图中, 有尖峰(相对于可行解空间中的局部极大值), 有低谷(相对于可行解空间中的局部极小值), 等等. step1 比 step2 大, 说明组 1 可行解空间比组 2 可行解空间显得更加“崎岖”, 在算法的搜索过程中, 就需要花费更多的时间才能确定一个解对于搜索过程所提供的有用信息, 相对来说, 组 1 的实例比组 2 的实例更难一些. 因此, 各工序加工时间的分布, 将影响到可行解空间中局部极小值点的多少; 时间分布越分散, 局部极小值点就越多, 从而使得调度问题就越复杂, 算法求解过程也就越困难.

5 小结

本文以 JSP 问题为例, 研究了调度问题本身的结构特性, 分析了调度问题可行解空间的属性, 提出了分割因子的概念. 研究表明, 分割因子影响调度问题可行解空间的规模, 而各工序加工时间的长短则影响局部极值点的分布情况. 实际上, 依照分割因子的概念, 也可以对工件加工时间的分布提出各种各样的因子, 从而在不同的层次上来表示可行解空间的属性. 有了这些因子, 就可以归纳总结不同的近优算法对不同的调度问题所表现出的性能和效率上的差异, 从中找出算法各个调整参数(如 GA 中的交叉率和 TA 中的禁忌列表长度等)与各因子的内在联系, 这样, 在运行调度算法之前对算法参数进行相应的调整, 以期获得满意的效果.

参考文献:

- [1] Christopher R Houck, Jeffrey A Jones, Michael G Kay. Characterizing search spaces for tabu search, currently under second review [J]. European Journal of Operational Research, 1999.
- [2] Dirk C Mattfeld, Christian Bierwith. A search space analysis of the job shop scheduling problem [J]. Baltzer Journals, April, 1996.
- [3] J Christopher Beck, W Ken Jackson. Constrainedness and the phase transition in job shop scheduling [M]. April 16, 1997.
- [4] E Weinberger. Correlated and uncorrelated fitness landscape and how to tell the difference [J]. Biol. Cybern 63, 1990.
- [5] Peter F Stsdlar. The landscape of traveling salesman problem [J]. Physics Letters A 161, 1992.
- [6] Eugeniusz Nowichi, Czeslaw Smutnichi. A fast taboo search algorithm for the job shop problem [J]. Management Science, June 1996, 42(1).
- [7] Peter JM. Van Laarhoven, et al. Job shop scheduling by simulated annealing [J]. Operations Research, January/February 1992, 40(1).
- [8] Robert H Sores, et al. New search spaces for sequencing problems with application to job shop scheduling [J]. Management Science, October 1992, 38(10).
- [9] Peter Buker. Scheduling Algorithm [M]. Springer-Verlag, Berlin, Heidelberg, 1998 (2nd).

作者简介:



孙元凯 男. 1972 年生于河北. 1991 年 9 月进入清华大学自动化系学习, 1996 年 7 月获得学士学位. 1996 年 9 月进入清华大学自动化系, 在国家 CIMS 工程研究中心攻读博士学位. 研究方向主要是车间生产过程的建模、调度和优化控制.

刘 民 1965 年生, 清华大学自动化系国家 CIMS 工程技术研究中心讲师、博士, 现为中国自动化学会名词委员会副主任委员, 已在国内外刊物和会议上发表学术论文 20 多篇, 目前感兴趣的领域为复杂制造系统智能优化理论与方法、人工生命、进化计算等.

吴 澄 中国工程院院士, 国家 863 计划自动化领域首席科学家, 清华大学自动化系教授, 博士生导师, 国家 CIMS 工程技术研究中心主任. 复杂制造系统智能优化理论与方法、人工生命、进化计算等.