

文章编号: 1001-0920(2009)12-1816-05

利用多变量自适应回归样条函数确定 ATCS 复合分派规则的缩放参数

陈艳^a, 杨菁^a, 李会敏^b

(澳门科技大学 a. 行政与管理学院, b. 基础教学部, 澳门)

摘要: 提出一种利用多变量自适应回归样条函数(MARS)来确定考虑准备时间的直观延误成本(ATCS)复合分派规则中缩放参数的方法, 以优化 ATCS 分派规则在最小化总加权延误时间(TWT)上的效果. 通过利用 MARS 模型在高维空间上的弹性建模能力, 构建调度作业组与缩放参数之间的非线性模型, 以便灵活地捕捉更多的局部映射关系. 对比实验结果表明, 与已有方法相比, 该方法可显著地改善 ATCS 规则在最小化总加权延误时间上的效果, 同时降低调度效果的不稳定性.

关键词: 多变量自适应回归样条函数; 考虑准备时间的直观延误成本; 分派规则; 缩放参数

中图分类号: F273

文献标识码: A

Using MARS to determine scaling parameter values for composite dispatching rule ATCS

CHEN Yan^a, YANG Jing^a, LI Hui-min^b

(a. Faculty of Administration and Management, b. General Studying, Macao University of Science and Technology, Macao, China. Correspondent: CHEN Yan, E-mail: yachen@must.edu.mo)

Abstract: A multivariate adaptive regression splines (MARS) based method is proposed to determine appropriate scaling parameter values for composite dispatching rule ATCS to generate good schedules, which aims at minimizing the total weighted tardiness. With the flexible piecewise structure in high-dimensional space, the MARS based model is able to reflect the local nonlinear relationship between the scaling parameter values and scheduling problem instances. Computational result shows that the proposed method outperforms the existing method in the literature in terms of scheduling result on total weighted tardiness and its variation.

Key words: Multivariate adaptive regression splines; Apparent tardiness cost with setups; Composite dispatching rules; Scaling parameters

1 引言

调度问题(Scheduling)是制造业和服务业在实践中常常需要面对的决策问题. 这类决策问题涉及到将有限的资源(如机器、人力以及空间等)分配给等待的作业、顾客或任务, 并产生某一个时间段内的工作计划表, 说明在每个时间点, 每种资源具体分配给哪一项任务^[1]. 大部分的调度问题都很难找到严格意义上的最优解, 即使是形式简单、没有太多约束条件的调度问题, 都有可能是强 NP-hard 的优化问题. 例如, 求解单台机器最小化总加权延误时间的 $1 \parallel \sum W_j T_j$ 调度问题^[2]. 因此, 在实践中, 当调度问题规模较大, 决策涉及到的工作数目较多时, 人们常使

用启发式算法迅速地求得一个较好的满意解. 分派规则便是启发式算法中的一种. 最常见的基本分派规则包括最早交付期(EDD)规则和加权最短处理时间(WSPT)规则. 分派规则依据各个等待被处理的工作的特性计算每个工作的优先等级, 并按照各项工作的优先等级来确定它们被处理的顺序. 例如, EDD 规则优先处理交付期早的工作; 而 WSPT 规则优先处理加权加工时间最短的工作. 分派规则可以在现实可行的运算时间里对调度问题给出一个较好的可行解.

实际中, 调度问题常常需要在多个目标之间进行权衡, 因此, 可将 EDD 这样只考虑单目标的基本

收稿日期: 2009-01-09; 修回日期: 2009-06-24.

作者简介: 陈艳(1974—), 女, 山东嘉祥人, 博士, 从事管理科学的研究; 杨菁(1976—), 女, 长沙人, 博士, 从事管理经济学的研究.

分派规则组合起来,以构建复合分派规则来同时优化多个目标.常见的复合分派规则有直观延误成本法(ATC)^[3]和考虑准备时间的直观延误成本(ATCS)规则^[4,5].

在利用基本分派规则构建复合分派规则时,每个基本分派规则都有自己的缩放参数,以决定其相对重要性.本文将讨论如何利用多变量自适应回归样条函数(MARS)来确定 ATCS 复合分派规则中的缩放参数的大小,以优化 ATCS 分派规则在最小化总加权延误时间(TWT)上的效果,并将该方法与文献中已有方法进行比较,以验证其有效性.

2 问题描述

首先, $Pm | s_{jk} | \sum \omega_j T_j$ 调度问题可表述为:有 n 项待加工的作业 $j = 1, 2, \dots, n$, 有 $i = 1, 2, \dots, m$ 台相同的平行机器, 记为 Pm . 各项作业只需在其中的一台机器上进行一次加工, 且每台机器每次只能处理一件工作. 工作 j 所需的加工时间记为 p_j , 交付期记为 d_j , 重要性记为 ω_j . 机器在连续加工工作 j 和 k 之间, 需要一定的准备时间 s_{jk} , 其大小取决于工作 j 和工作 k . 工作 j 的完成时间记为 C_j , 延误时间记为 $T_j = \max(C_j - d_j, 0)$. 该问题的求解目标是 minimized 作业的总加权延误时间. 根据 Lawler 等^[6] 提出的调度问题的 $\alpha | \beta | \gamma$ 标准表达形式, 该问题可表示为 $Pm | s_{jk} | \sum \omega_j T_j$, 这是一个 NP-hard 的问题.

1997 年, Lee 等^[5,6] 在直观延误成本(ATC)规则的基础上, 提出了考虑作业间准备时间的直观延误成本(ATCS)规则, 用于最小化有作业间准备时间的调度问题的总加权延误时间 TWT. 在该规则中, 每个等待被处理的工作的优先级按下式计算:

$$I_{ATCS}(t, l) = \frac{\omega_j}{p_j} \exp\left(-\frac{\max(d_j - p_j - t, 0)}{k_1 \bar{p}}\right) \exp\left(-\frac{s_{lj}}{k_2 \bar{s}}\right). \quad (1)$$

式中: I 代表优先级, t 表示当前的时间, l 表示上一个在该机器上被处理的工作, \bar{p} 表示平均加工时间, \bar{s} 表示平均的作业间准备时间. 每当有机空闲时, 具有最大优先级的工作被选中加工.

ATCS 复合分派规则实际上是加权最短处理时间规则、最小宽松时间(LSR)规则和最小准备时间(SST)规则的组合. 在式(1)中, k_1 和 k_2 为 ATCS 复合分派规则的缩放参数. 缩放参数 k_1 用来确定 $\exp\left(-\frac{\max(d_j - p_j - t, 0)}{k_1 \bar{p}}\right)$ 项, 即 LSR 规则的权重; k_2 用来确定 $\exp\left(-\frac{s_{lj}}{k_2 \bar{s}}\right)$ 项, 即 SST 规则的权重; 同时, 该等式有一个隐含的 $k_0 = 1$, 决定了 ω_j/p_j 这一项, 即 WSPT 规则的权重. 显然, 缩放参数 k_1 和 k_2

的取值不同, 会导致不同的作业处理次序, 影响 ATCS 规则在最小化总加权延误时间 TWT 上的效果.

3 研究现状

目前文献中关于复合分派规则的讨论大多侧重于规则的构造, 即如何将考虑单目标的基本分派规则进行数学表达式上的组合以达到综合优化多个调度目标的目的. 其中, 缩放参数常被用来在构造规则时在多个目标之间进行权衡. 目前已有文献关于缩放参数值的确定大致有以下 3 类方法.

第 1 类方法是使用经验值. Rachamadugu 和 Morton^[7] 根据经验建议如下复合分派规则 RM:

$$I_{RM}(t) = \frac{\omega_j}{p_j} \exp\left(-\frac{\max(d_j - p_j - t, 0)}{kp}\right), \quad (2)$$

其缩放参数 k 的值应在 0.5 到 2 之间.

Vepsalainen 和 Morton^[3] 建议在使用如下 ATC 规则时:

$$I_{ATC}(t) = \frac{\omega_j}{p_j} \exp\left[-\frac{\max\left\{d_j - p_j - t - \sum_{q=i+1}^{m_j} (\omega_{qj} + p_{qj}), 0\right\}}{kp}\right], \quad (3)$$

如果是静态的流程型作业的调度, 则缩放参数 k 可取 2; 如果是动态的零工型作业的调度, 则缩放参数 k 可取 3.

第 2 类方法是根据作业组的特性来确定缩放参数的取值. Morton 和 Pentico^[8] 使用等式 $B = 1.3 + \rho$ 来确定如下 X-RM 复合分派规则的缩放参数 B 的值:

$$I_{X-RM}(t) = \frac{\omega_j}{p_j} \exp\left(-\frac{\max(d_j - p_j - t, 0)}{kp}\right) \times \left(1 - \frac{B}{P_{\min}} \max(r_j - t, 0)\right). \quad (4)$$

式中: ρ 表示机器的利用率, r_j 表示作业 j 的可开始操作时间, P_{\min} 表示所有等待作业的最小加工时间.

Lee 等^[4,5] 使用曲线拟合的方法来估计 ATCS 规则的缩放参数 k_1 和 k_2 的值. 利用作业组的 4 个特性因子: 作业机器比例 μ , 准备时间重要性 η , 交付期紧张度 τ 和交付期分散度 R , 则缩放参数 k_1 和 k_2 的值可由以下拟合函数进行估计:

$$k_1 = 1.2 \ln(\mu) - R, \quad k_2 = \frac{\tau}{A_2 \sqrt{\eta}}. \quad (5)$$

当 $\tau < 0.8$ 时, $A_2 = 1.8$; 当 $\tau \geq 0.8$ 时, $A_2 = 2.0$. 同时, k_1 的值需作以下调整:

- 1) 如果 $\tau < 0.5$, 则 $k_1 = k_1 - 0.5$;
- 2) 如果 $\eta < 0.5$ 且 $\mu > 5$, 则 $k_1 = k_1 - 0.5$.

文献[9,10]利用回归模型 $k = F(\mu, \tau, \eta, R, J_a, r_{-\tau})$ 来估计 ATCSR(Apparent tardiness cost with setups and ready times) 规则里缩放参数 k_1, k_2 和 k_3 的值. 其中: J_a 为可立即加工工作占所有等待工作的百分比, $r_{-\tau}$ 为可开始加工时间分散度因子.

第 3 类方法是利用后启发式算法来确定缩放参数的取值. Kim 等^[11] 利用 μ, τ, η, R 作为输入节点, 使用反向传播的人工神经网络算法为 ATC 和 ATCS 规则估计缩放参数的值. Grabot 等^[12] 讨论和比较了如何利用专家系统和人工神经网络来确定模糊分派规则里的权重参数.

1991 年, Friedman^[13] 首次提出将 MARS 方法用于高维预测变量 ($3 \leq$ 预测变量数目 ≤ 20) 的弹性映射模型的建立. MARS 在由多维预测变量和响应变量构成的高维空间里, 以分段多项式函数的形式映射预测变量与响应变量之间的关系. 因此, 与构建一个单一的全局的方程相比, MARS 有可能捕捉到更多的真实映射关系的区域特性.

4 方法描述

4.1 数据收集

Pm | s_{jk} | $\Sigma w_j T_j$ 调度问题^[5], 有 4 个用以描述作业组特征的因子:

1) 作业机器比例 $\mu = n/m$, 用来估计每台机器平均要处理多少个工作.

2) 准备时间重要性 $\eta = \bar{s}/\bar{p}$, 用来衡量准备时间相对于整个调度的重要程度.

3) 交付期紧张度 $\tau = 1 - (\bar{d}/C_{\max})$ ^[5], 衡量的是交付期的紧张程度. 当 τ 接近 1 时, 表明工期紧张; 当 τ 接近 0 时, 表明工期松弛.

4) 交付期分散度 $R = (d_{\max} - d_{\min})/C_{\max}$ ^[5], 衡量的是各项工作的交付期的分散程度.

以下所研究的作业组特性的变化范围与文献[5]中一致. 具体区域如下: $\mu \in [5, 30]$, $\eta \in [0.25, 1.0]$, $\tau \in [0.3, 0.9]$ 和 $R \in [0.25, 1.0]$. 该区间涵盖了绝大部分对实际中的调度问题有指导意义的变化范围, 也可将本文方法与文献中的既有结果做一比较. 如果实际中要调度的作业组特性不处于此研究区域, 则本文方法仍然适用, 只是得出的映射模型会有所区别.

首先, 在如下的 560 个不同特性因子的组合上随机生成 10 个具有该特性因子水平的作业组, 共 5600 个. 作业组的具体生成方法见文献[5].

$$\mu = (5, 10, 15, 20, 30),$$

$$\tau = (0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9),$$

$$\eta = (0.25, 0.5, 0.75, 1.0),$$

$$R = (0.25, 0.5, 0.75, 1.0).$$

对于每个作业组, 利用格点搜索法, 在离散区间 $k_1 = (0.2, 0.4, 0.6, \dots, 6.4)$ 和 $k_2 = (0.1, 0.2, 0.3, \dots, 1.6)$ 内搜索使总加权延误时间 TWT 最小的缩放参数组合 (k_1, k_2) , 并记录下来. 最终得到 5600 组 6 维数据. 4 维自变量分别是作业机器比例、准备时间重要性、交付期紧张度和交付期分散度; 2 维响应变量分别是使总加权延误时间最小的缩放参数 k_1 和 k_2 .

4.2 模型建立

利用上节得到的 5600 组数据, 使用统计软件 R (version 2.2.1) by R Development Core Team^[13] 构建 MARS 模型, 即

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X). \quad (6)$$

式中: 系数 β 利用最小二乘法来估计, $h_m(X)$ 是一个基函数或两个基函数之积. 基函数的构造如下: 在输入, 即预测变量每间隔 $1/3$ 单位处取一个节点 t , $(x-t)^+$ 和 $(t-x)^+$ 称为一个基函数对. 所有基函数构成的集合为基函数集, 记为 C , 即

$$C = \{(x_j - t)^+, (t - x_j)^+\}_{t \in \{x_{11}, x_{12}, \dots, x_{4m}\}} \quad (7)$$

通过允许 $h_m(X)$ 是两个基函数之积, 实际上在建模过程中考虑了两个自变量的交互作用. 模型的建立使用了前向逐步回归法. 该建模策略从一个常数函数 $h_0(X) = 1$ 开始, 然后每次向模型增加新的一项

$$\hat{\beta}_{M+1} h_l(X) \cdot (x_j - t)^+ + \hat{\beta}_{M+2} h_l(X) \cdot (t - x_j)^+.$$

新项的构造和选择方法如下: $h_l(X)$ 是任何一个已被包含在模型里的函数, $(x-t)^+$ 和 $(t-x)^+$ 是基函数集里的一个基函数对. 增加的 $\hat{\beta}_{M+1} h_l(X) \cdot (x_j - t)^+ + \hat{\beta}_{M+2} h_l(X) \cdot (t - x_j)^+$ 项, 是所有

$$\hat{\beta}_{M+1} h_l(X) \cdot (x_j - t)^+ + \hat{\beta}_{M+2} h_l(X) \cdot (t - x_j)^+$$

里能最大地减少残差平方和 (SSR) 的组合. 重复这样的步骤, 直至模型所含的项数达到预设的最大值. 软件 MARS by Salford Systems^[15] 的用户手册建议, 该最大值应设置成“真实”函数关系所含项数的 2~4 倍. 在本文中, 最大值设为 400, 以保证建模能充分展开. 此后, 进行后向的淘汰过程. 每一步, 经删减之后使得模型的残差平方和 SSR 增加最小的一项, 直到模型仅剩下常数函数 $h_0(X) = 1$. 然后, 计算每个中间生成模型的 GCV (Generalized cross validation) 指标值, 即

$$\text{GCV} = \frac{\sum_{i=1}^N (y_i - \hat{f}(x_i))^2}{(1 - M(\lambda)/N)^2}. \quad (8)$$

GCV 最小的模型被选作最终的最好模型.

式(8)中: $M(\lambda)$ 表示模型的复杂程度, N 是用以建模数据的个数. 函数 $M(\lambda)$ 决定模型为每个包

含的基函数和节点而丧失多少的自由度；参数 λ 的值通过交叉验证来确定，本文通过交叉验证，在缩放参数 k_1 和 k_2 的建模中， λ 分别取为 0.2 和 1.

得到的 MARS 模型如下所示 (R^2 分别是 59.8% 和 60.6%)：

$$\begin{aligned}
 k_1 = & 1.61 - 0.05\mu - 2.77\eta + 0.27\tau + \\
 & 8.78(\eta - 0.75)^+ - 2.98(\eta - 0.23)^+ - \\
 & 9.13R + 0.71(R - 0.19)^+ + 0.25\mu R - \\
 & 0.22\mu(R - 0.19)^+ + 0.29\mu\tau - 0.20\mu(\tau - \\
 & 0.75)^+ + 6.40\tau R + 4.66(\tau - 0.48)^+ - \\
 & 0.29\mu(\tau - 0.48)^+ + 6.10\eta R - 4.49(\tau - \\
 & 0.85)^+ + 10.36(R - 0.10)^+ + 20.27(\eta - \\
 & 1.00)^+ - 6.31(\tau - 0.71)^+ + \\
 & 0.1\mu\eta - 7.04(R - 0.07)^+ - 5.90(\tau - 0.96)^+ - \\
 & 0.09(R - 0.15)^+ - 5.14\eta(R - 0.15)^+ - \\
 & 4.98(\tau - 0.48)^+ R + 5.66(R - 0.65)^+ - \\
 & 4.15\eta(R - 0.65)^+ + 10.61\eta\tau - 10.70\eta(\tau - \\
 & 0.23)^+ + 1.06(R - 0.57)^+ - 0.20\mu(R - \\
 & 0.57)^+ + 1.1(R - 0.44)^+ + 3.08(\tau - \\
 & 0.96)^+ R - 0.87(\tau - 0.55)^+, \tag{9}
 \end{aligned}$$

$$\begin{aligned}
 k_2 = & 0.30 - 0.36\eta + 1.64\tau + 0.33(\eta - \\
 & 0.50)^+ - 1.54\eta\tau + 1.41(\eta - 0.63)^+ - \\
 & 0.61R + 0.68(\eta - 0.50)^+ \tau + \\
 & 1.87(\eta - 0.99)^+ + 0.89\tau R - 3.59(\tau - \\
 & 0.63)^+ R - 1.00(\tau - 0.71)^+ + 0.17\eta R + \\
 & 0.10(\eta - 0.75)^+ + 2.39(\tau - 0.71)^+ R - \\
 & 0.17(R - 0.57)^+ + 2.14(\eta - 0.94)^+ - \\
 & 0.51(\tau - 0.48)^+ - 1.49(\tau - 0.65)^+ + \\
 & 0.72\eta(\tau - 0.65)^+ - 0.002\mu + \\
 & 0.004\mu R + 0.01\mu\eta - 0.003\mu\tau + \\
 & 0.03 \mu(\tau - 0.63)^+. \tag{10}
 \end{aligned}$$

表 1 和表 2 给出了对于 k_1 和 k_2 ，各因子自变量的重要程度排序。表中的“+ SSR”一栏给出的是建模过程中不引入该变量时所带来的 SSR 的增加；“重要性”一栏给出的是基于 + SSR 的值所估计的各预测变量的相对重要性，其值等于该预测变量的 + SSR 相对于最大 + SSR 的百分比值。显然，最重要

表 1 各自变量对于 k_1 的相对重要性

因子 / 预测变量	+ SSR	重要性 / %
1 μ	5138.87	100.00
2 τ	1990.26	38.73
3 R	348.88	6.79
4 η	259.41	5.05

表 2 各自变量对于 k_2 的相对重要性

因子 / 预测变量	+ SSR	重要性 / %
1 η	216.81	100.00
2 τ	115.75	53.39
3 R	2.12	0.98
4 μ	2.12	0.98

的预测变量是能最好解释缩放参数取值变化的因子。

对于 k_1 ，最显著的因子是作业机器比例 μ 和交付期紧张度 τ 。交付期分散度 R 和准备时间重要性 η 相对而言不是很重要。这个结论与文献[5]中的发现一致。同时，图 1 和图 2 中 k_1 的等值变化曲线与文献[5]中的大体类似，但与文献[5]相比，图 1 和图 2 灵活地显示出更多的不规则的局域性特征。

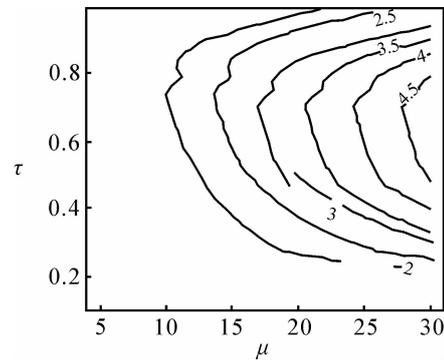


图 1 k_1 的 μ 和 τ 等值曲线

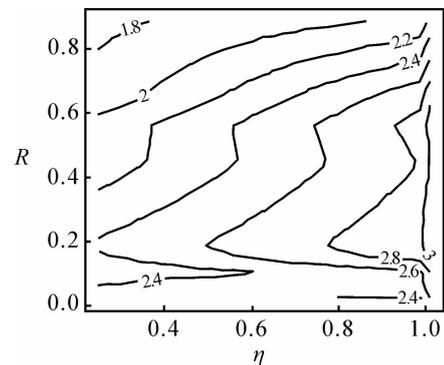


图 2 k_1 的 η 和 R 等值曲线

对于 k_2 ，最显著的因子是准备时间重要性 η 和工期紧张度 τ 。同样，这与文献[5]中的发现一致。

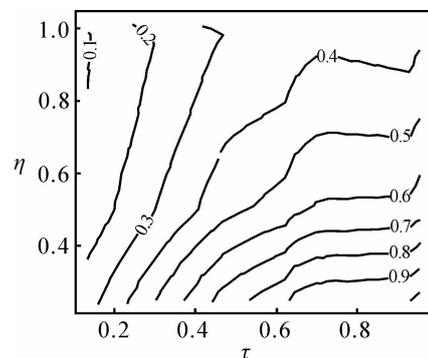


图 3 k_2 的 τ 和 η 等值曲线

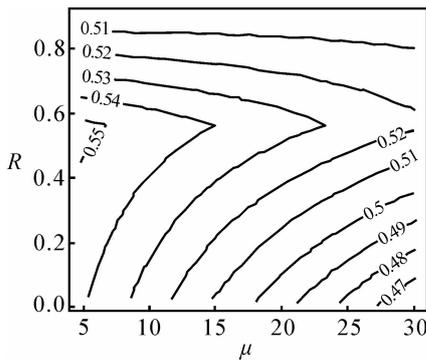


图 4 k_2 的 μ 和 R 等值曲线

图 3 和图 4 中 k_2 的等值变化曲线同样显示出了显著的局域特征。

5 模型有效性的验证

为了验证以上所得到的 MARS 模型,采用随机生成的新的作业组来检验其有效性.该模型被用来确定 ATCS 分派规则中 k_1 和 k_2 的值,之后该规则将对新生成的作业组进行调度,得到的总加权延误时间 TWT 与文献[5]中方法所得到的结果进行比较.文献[5]中的拟合曲线方法,首先在多个二维空间上描绘出作业组的特性与缩放参数的映射曲线,然后根据曲线的形状拟合成函数式(5).新的作业组由如下因子组合生成: $\mu - (6, 12, 20, 30), \eta - (0.35, 0.6, 0.9), \tau - (0.45, 0.75)$ 和 $R - (0.35, 0.6, 0.9)$, 每个组合随机生成 10 个作业组,共得到 720 个不同的作业组.

效果比率被用来评估 MARS 模型和拟合曲线的有效性,其值等于用该方法得到的总加权延误时间与用格点搜索法所得到的最小加权总延误时间之比,即

$$PR(k_1, k_2) = \frac{TWT(k_1, k_2)}{\min TWT(\text{GridSearch})}. \quad (11)$$

PR 的值越小,说明相对的加权总延误时间越小,该调度结果越好.

表 3 给出了两种方法得到的 720 个 PR 值的统计结果.表 3 显示,由拟合曲线方法得到的 PR 值在 0 ~ 101.371 之间变化,均值是 2.781,标准差是 5.593.由 MARS 方法得到的 PR 的值在 0 ~ 68.1920 之间变化,均值是 1.3740,标准差是 2.6282.数据显示,利用 MARS 方法来估计缩放参数的值可以很大程度地改善 ATCS 规则的效果,不仅平均的总加权延误时间小,而且差异性更小,表现更稳定.这两种方法的 PR 值的分布都向右倾斜,表明它们的均值容易受到极大值的影响,因此应进一步考查它们的中位数.由拟合曲线方法得到 PR 的中位数是 1.413,由 MARS 方法得到 PR 的中位数

是 1.0759. 这组数据同样显示, MARS 方法明显优于拟合曲线方法.

表 3 MARS 模型和拟合曲线方法的调度结果比较

参数确定	表现比率					
	方法	最小值	最大值	均值	中位数	标准差
拟合曲线		0	101.371	2.781	1.413	5.593
MARS 模型		0	68.1920	1.3740	1.0759	2.6282

进一步,在同一个坐标系里利用箱线图观察由这两种方法得到的 PR 值的分布情况.图 5 显示, MARS 方法得到的 PR 值所集中的区间要明显低于(即小于)拟合曲线方法得到的 PR 的区间,在这个具体的问题中,说明用 MARS 方法得到的总加权延误时间更小.

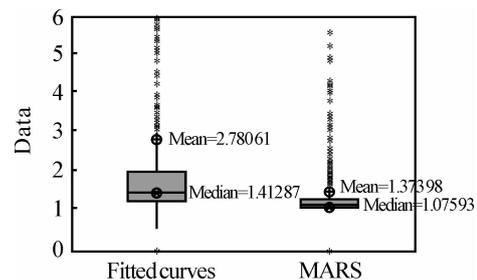


图 5 PR 值的箱线图分布比较

6 结 论

本文提出了一种利用多变量自适应回归样条函数 MARS 确定复合分派规则 ATCS 的缩放参数的方法,并将它与文献中现有的拟合曲线方法进行了比较.通过对随机生成的样本作业组进行检验,结果显示, MARS 的方法可以显著改善 ATCS 规则在最小化总加权延误时间的效果,同时降低调度效果的不稳定性.将此方法扩展到其他复合分派规则,如 ATC, ATCSR 和其他的参数设定问题,是进一步研究的方向.

参考文献(References)

[1] Pinedo M. Scheduling: Theory, algorithms and systems [M]. 2nd ed. New Jersey: Prentice-Hall, 2002.
 [2] Du J, Leung J Y. Minimizing total tardiness on one machine is NP-hard [J]. Mathematics of Operations Research, 1990, 15(3): 483-494.
 [3] Vepsalainen A, Morton T. Priority rules for job shops with weighted tardiness costs [J]. Management Science, 1987, 33(8): 1035-1047.
 [4] Lee Y H, Bhaskaran K, Pinedo M. A heuristic to minimize the total weighted tardiness with sequence-dependent setups [J]. IIE Transactions, 1997, 29(11): 45-52.