动态电源管理的随机切换模型与在线优化

江 琦[1] 奚宏生[1] 殷保群[1]

摘 要 考虑系统参数未知情况下的动态电源管理问题，提出一种基于强化学习的在线策略优化算法. 通过建立事件驱动的随机切换分析模型，将动态电源管理问题转化为带约束的 Markov 决策过程的策略优化问题. 利用此模型的动态结构特性，结合在线学习估计梯度与随机逼近改进策略，提出动态电源管理策略的在线优化算法. 随机切换模型对电源管理系统的动态特性描述精确，在线优化算法自适应性强，运算量小，精度高，具有较高的实际应用价值.

关键词 动态电源管理，Markov 决策过程，强化学习，梯度估计，随机逼近，在线优化
中图分类号 TP202

Stochastic Switching Model and Policy Optimization Online for Dynamic Power Management

JIANG Qi XI Hong-Sheng YIN Bao-Qun

Abstract A reinforcement learning based online optimization algorithm is presented for dynamic power management with unknown system parameters. First an event-driven stochastic switching model is introduced to formulate dynamic power management problem as a constrained policy optimization problem. Then by utilizing the features of this model an online optimization algorithm that combines policy gradient estimation and stochastic approximation is derived. The stochastic switching model captures the power-managed system behaves accurately. The optimization algorithm is adaptive, and can achieve global optimum with less computational cost. Simulation results demonstrate the effectiveness of the proposed approach.

Key words Dynamic power management, Markov decision processes, reinforcement learning, gradient estimation, stochastic approximation, on-line optimization

1 引言

动态电源管理是一种系统级的功耗控制技术，广泛应用于便携式电子装置、移动通信终端以及网络设备的功耗控制. 在实际使用中，系统组件的工作负荷随时间动态变化，动态电源管理通过将负荷较轻的组件切换到较低功耗 (对应于较低的性能) 的运行状态，在满足性能要求的同时，降低系统的功耗. 功耗控制的效果取决于动态电源管理策略的优劣，其控制策略的选取是一个在性能约束下的最小功耗的带约束优化问题.

通常采用的动态电源管理策略有三种类型: 1) "Time-out" 策略[1], 即将系统组件在空闲设定的时间间隔 (阈值) 后切换到低功耗状态. 2) 预测式策略[2], 基于工作负荷的相关性假设，当电源管理控制器预测系统组件的下一个空闲周期大于设定值时，将系统组件在转为空闲的时刻切换到低功耗的状态. 这两类策略都属于启发式的，不能保证最佳的应用效果，且局限于只有两种运行状态的应用. 3) 随机型策略，通过建立随机模型来描述电源管理系统，并采用随机优化方法进行策略优化，应用的效果取决于系统模型的精确性和优化算法的优劣. 文献 [3,4] 建立了离散时间 Markov 决策过程模型，运用线性规划的方法进行策略优化，由于需要在每个离散时刻进行策略评估，计算量较大. 文献 [5] 建立了基于连续时间 Markov 决策过程的系统模型，运用改进的策略迭代算法求解最优策略，其模型没有精确描述系统处于运行状态转移过程中的特征，策略优化局限于确定型策略，算法在迭代过程中需要不断调整权重系数的取值，增加了计算的复杂度.

动态电源管理系统由于应用环境的复杂性，系统参数难以预先精确获取且具有时变的特点. 本文通过建立动态电源管理系统的事件驱动的随机切换模型，利用此模型的动态结构特性[6], 提出一种基于策略梯度的在线学习和优化算法. 该算法不依赖于系统参数的信息，具有较强的自适应性；无需计算各状态的性能势或其他相关量 (如 Q -因子), 有效减

收稿日期 2005-10-20 收修改稿日期 2006-7-15
Received October 20, 2005; in revised form July 15, 2006
国家自然科学基金 (60574065), 国家 863 计划 (2005AA103320), 安徽省自然科学基金 (050420301) 资助
Supported by National Natural Science Foundation of P. R. China (60574065), National 863 Program of P. R. China (2005AA103320), and Anhui Provincial Natural Science Foundation of P. R. China (050420301)
1. 中国科学技术大学自动化系 合肥 230027
1. Department of Automation, University of Science and Technology of China, Hefei, 230027
DOI: 10.1360/aas-007-0066

少计算量，提高实时性；能够确保收敛到全局最优，克服了策略梯度法通常只能收敛到局部最优的固有缺陷. 该模型与算法的有效性通过一个具体应用的仿真试验加以验证.

2　系统模型

2.1　动态电源管理问题

系统级的动态电源管理通过有选择地将处于空闲 (或较低工作负荷) 的系统组件切换到较低功耗的状态，从而有效降低系统的功耗. 可管理电源组件能够提供多种运行状态，对应不同的工作性能和功率消耗, 根据电源管理控制器的控制指令，进行运行状态的切换. 将系统组件切换到低功耗运行状态的同时，也相应降低了系统组件工作性能，同时运行状态的切换需要一定的电能消耗和一定的转换时间，使得功耗增加和性能降低. 不适当的切换非但不能带来功耗的降低，而且导致较大的性能损失. 动态电源管理在系统的性能和功耗间进行均衡，策略优化即寻求一种达到边界最优的控制策略.

动态电源管理系统由等待服务的队列 SQ (Service queue)、服务处理器 SP (Service provider)、电源管理控制器 PM (Power manager) 组成. 服务请求 SR (Service request) 的到达由系统所处的环境决定. 系统提供一定容量的等待队列 SQ，存储未能及时得到处理的服务请求. SP 具有多种运行状态，分别对应于不同的服务率和功耗，按照 PM 的控制指令进行运行状态的切换. PM 根据系统的运行情况和控制策略，发布控制指令，将 SP 切换到合适的运行状态. 动态电源管理问题即寻找一种最优控制策略, 对 SP 的运行状态进行切换控制，目标是使系统在满足性能要求的同时，功率消耗最小.

实际应用中, 系统的服务请求到达的时间间隔、所需的处理时间、运行状态转换时间具有随机的概率分布，当满足或近似满足指数分布时，可以构建动态电源管理问题的 Markov 决策过程模型进行性能分析与策略优化. 动态电源管理系统所处的应用环境复杂，服务到达率甚至服务率往往是时变的参数，使得构建的系统模型是非时齐的，增加了问题处理的复杂性. 这里通过将时变的服务到达率即环境的变化描述为一个连续时间 Markov 过程，并假定对于特定种类的服务请求，SP 各运行状态的服务率是恒定不变的，从而将这一非时齐问题转化为一个时齐问题来讨论，其结果可进一步推广至一般时变参数情况下的应用，如根据环境的变化动态调整随机逼近算法的步长来增强算法对时变参数的适应性.

2.2　随机切换模型

考虑处于某种工作环境中的动态电源管理系统，为单一种类的服务请求提供服务. 服务请求到达可以用独立的 Possion 过程来描述，到达率为 λ_r，其中 r 表示系统的环境状态，r 所有可能的取值构成系统的环境状态空间 S_R，环境的变化用一个连续时间 Markov 过程 $\mathrm{M}_R = \{S_R, A_R\}$ 来表示，其状态转移速率矩阵 $A_R = [a_{rr'}], r, r' \in S_R$.

SQ 容量为 n_Q，排队规则为 FIFO, 以处于系统中的服务请求的个数 q 表示系统的内部状态，内部状态空间为 S_Q.

SP 的运行状态以 p 表示，服务处理时间服从指数分布，服务率和功率消耗分别为 μ_p、c_p，S_P 为运行状态的集合. 运行状态之间的转换过程以 k 表示，转换过程所需时间服从指数分布，平均转换时间、服务率和功率消耗分别以 τ_k、μ_k、c_k 表示，S_K 为运行状态转换过程的集合. 以 $S_S = S_P \cup S_K$ 表示扩展的运行状态空间，s 表示扩展的运行状态.

PM 根据控制策略选取切换行动，将 SP 切换至适当的运行状态. 切换行动空间为 $D = \{d_{ss'}, s, s' \in S_S\}$，其中行动 $d_{ss'}$ 表示将运行状态从 s 切换至 s'. 以 $\theta_{ss'}^{rq}$ 表示系统处于环境状态 r，内部状态转移至 q 时选择行动 $d_{ss'}$ 的概率. PM 的控制策略由 $\boldsymbol{\theta} = (\theta_{ss'}^{rq}, r \in S_R, s, s' \in S_S, q \in S_Q)^{\mathrm{T}}$ 确定，$\boldsymbol{\theta}$ 所有可能的取值构成随机 Markov 型策略集合的参数化形式 $\Theta = \{(\theta_{ss'}^{rq}, r \in S_R, s, s' \in S_S, q \in S_Q)^{\mathrm{T}} | 0 \le \theta_{ss'}^{rq} \le 1, \sum_{s' \in S_S} \theta_{ss'}^{rq} = 1\}$.

对应于固定环境状态 r，运行状态 s，系统为一个 M/M/1 排队系统，运行规律可以用一个有限状态的连续时间 Markov 过程 M_Q^{rs} 来表示，其状态空间为 S_Q，转移速率矩阵 $A_Q^{rs} = [a_{qq'}^{rs}]$. 对应固定的环境状态 r，系统为有限个 $\mathrm{M}_Q^{rs}, s \in S_S$ 通过切换控制策略 $\boldsymbol{\theta} = (\theta_{ss'}^{rq}, r \in S_R, s, s' \in S_S, q \in S_Q)^{\mathrm{T}}$ 构成的 Markov 切换控制过程 M_{SQ}^r，其状态空间为 $S_{SQ}^r = S_S \times S_Q$，转移速率矩阵为 $A_{SQ}^r(\boldsymbol{\theta})$. 考虑环境状态的变化，因 M_R 与 M_{SQ}^r 相互独立，系统可以综合成一个 Markov 过程 $\mathrm{M} = \{X_t, t \ge 0\}$，其状态空间为 $S = S_R \times S_{SQ}$，状态转移矩阵 $A(\boldsymbol{\theta})$，其中的元素为:

$$
a_{(rsq)(r's'q')}(\boldsymbol{\theta}) = \begin{cases} a_{qq'}^{rs} + a_{rr'} - a_{ss''}, & r = r', s = s', q = q' \\ a_{qq'}^{rs} \cdot \theta_{s'}^{rq'}, & r = r', q \ne q' \\ a_{ss'}, & r = r', s \ne s', q = q' \\ a_{rr'}, & r \ne r', s = s', q = q' \\ 0, & \text{其他} \end{cases} \tag{1}
$$

式中 $a_{ss'}(a_{ss''})$ 为运行转换状态至目标状态的转移速率，当 $s \in S_K$，并且其转换目标运行状态为 $s'(s'') \in S_P$ 时，$a_{ss'}(a_{ss''}) = 1/\tau_s$，其他情况

$a_{ss'}(a_{ss''}) = 0$.

设性能函数 $f_p : S \to \mathcal{R}$，功耗函数 $f_c : S \to \mathcal{R}$，$f_c(rsq) = c_s$，对应于各状态的功率消耗. 定义系统的平均性能测度与平均功耗测度分别为

$$\eta_p(\boldsymbol{\theta}) = \lim_{T\to\infty}\frac{1}{T}\mathrm{E}\Big[\int_0^T f_p(X_t)\mathrm{d}t\Big]$$

$$\eta_c(\boldsymbol{\theta}) = \lim_{T\to\infty}\frac{1}{T}\mathrm{E}\Big[\int_0^T f_c(X_t)\mathrm{d}t\Big]$$

动态电源管理系统的随机切换模型为

$$\{S,\ A(\boldsymbol{\theta}),\ D,\ (f_p, f_c),\ (\eta_p(\boldsymbol{\theta}), \eta_c(\boldsymbol{\theta}))\}$$

动态电源管理问题表示成一个带约束的优化问题，即寻找一个最优策略 $\boldsymbol{\theta}^*$，使得在满足性能要求 G 的条件下系统的功耗最小

$$\begin{aligned}&\mathrm{PO}:\ \min_{\theta\in\Theta}\eta_c(\boldsymbol{\theta})\\ &\text{s.t.}\ \ \eta_p(\boldsymbol{\theta}) \geq G\end{aligned}$$

3 在线策略优化

在线优化是指在系统运行过程中不断进行策略改进，随机逼近最优策略. 以基于性能势 (即 relative cost vector[7]或 bias[8], 它们之间相差一个常数) 的性能梯度公式和性能差公式为基础, 根据优化的策略空间类型的不同，在线学习与优化方法可分为两类[9]: 基于值迭代和策略迭代的如 Q -学习方法和在线策略迭代方法[10], 适用于优化确定型策略; 基于策略梯度的如在线梯度估计方法[11]、PA 方法[12]、相似率方法[13]，适用于优化随机型策略. 动态电源管理问题是一个带约束的策略优化问题，其最优策略属于随机型策略，当系统参数未知时，可以运用策略梯度法进行在线学习与优化.

3.1 在线学习估计性能梯度

连续时间 Markov 过程 $\mathrm{M} = \{X_t, t \geq 0\}$ 的状态空间 $S = \{1, 2, \cdots, N\}$ 有限，转移速率矩阵 $A(\boldsymbol{\theta}) = [a_{ij}(\boldsymbol{\theta})]$, $i, j \in S$ 是参数向量 $\boldsymbol{\theta} \in \Theta$ 的函数. 由 (1) 式有，$a_{ij}(\boldsymbol{\theta})$, $i, j \in S$ 有界、一阶导数有界及二次可导，对于 $\forall\, \boldsymbol{\theta} \in \Theta$, $A(\boldsymbol{\theta})$ 不可约[8]，M 是遍历的 Markov 过程，存在与初始状态无关的唯一平稳分布 $\boldsymbol{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), p_2(\boldsymbol{\theta}), \cdots, p_N(\boldsymbol{\theta}))$，且满足

$$\boldsymbol{p}(\boldsymbol{\theta})\boldsymbol{e} = 1,\ \ \boldsymbol{p}(\boldsymbol{\theta})A(\boldsymbol{\theta}) = 0,\ \ A(\boldsymbol{\theta})\boldsymbol{e} = 0 \tag{2}$$

性能函数 f (f_p 或 f_c) 表示成向量形式 $\boldsymbol{f} = (f_1, f_2, \cdots, f_N)^{\mathrm{T}}$, 则平均性能测度$\eta(\boldsymbol{\theta}) = \boldsymbol{p}(\boldsymbol{\theta})\boldsymbol{f}$. 定义 Poisson 方程 $A(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{\theta}) = -\boldsymbol{f} + \eta(\boldsymbol{\theta})\boldsymbol{e}$ 的解 $\boldsymbol{g}(\boldsymbol{\theta})$ 为性能势向量，其第 i 个分量 $g_i(\boldsymbol{\theta})$ 为状态 $i \in S$ 的性能势[14].

引理 1.[14] 平均性能测度关于策略参数的导数

$$\nabla\eta(\boldsymbol{\theta}) = \sum_{i\in S} p_i(\boldsymbol{\theta}) \sum_{j\in S} \nabla a_{ij}(\boldsymbol{\theta}) \cdot g_j(\boldsymbol{\theta}) \tag{3}$$

由 (2) 式，有

$$\nabla a_{ii}(\boldsymbol{\theta}) = -\sum_{j\in S, j\neq i} \nabla a_{ij}(\boldsymbol{\theta})\ ,\ \ i \in S$$

记 $S_i = \{j | a_{ij} > 0\}$，则 (3) 式可写成

$$\begin{aligned}&\nabla\eta(\boldsymbol{\theta}) =\\ &\sum_{i\in S} p_i(\boldsymbol{\theta}) \sum_{j\in S, j\neq i} \nabla a_{ij}(\boldsymbol{\theta}) \cdot (g_j(\boldsymbol{\theta}) - g_i(\boldsymbol{\theta})) =\\ &\sum_{i\in S} p_i(\boldsymbol{\theta}) \sum_{j\in S_i} a_{ij}(\boldsymbol{\theta}) \cdot (g_j(\boldsymbol{\theta}) - g_i(\boldsymbol{\theta})) \cdot\\ &\nabla a_{ij}(\boldsymbol{\theta}) / a_{ij}(\boldsymbol{\theta})\end{aligned} \tag{4}$$

式中 $p_i(\boldsymbol{\theta}) \cdot a_{ij}(\boldsymbol{\theta})$ 有直观的物理意义：在稳态下，单位时间内，从状态 i 转移到 j 的次数.

设 $\{X_t,\ t \geq 0\}$ 是由 $A(\boldsymbol{\theta})$ 生成的一条样本轨道，$i^* \in S$ 是一个正常返状态，t_m 是 m 次抵达状态 i^* 的时间. $\{X_t,\ t \geq 0\}$ 在每一时刻 $t = t_m$ 后的延续在统计意义上以概率 1 等价，t_m 为再生时刻，i^* 为再生状态，X_t 是一个再生过程. $\{X_t,\ t_m \leq t < t_{m+1}\}$ 为第 m 个再生周期，周期的长度 $T_m = t_{m+1} - t_m$. t_m^n 表示在此周期中第 n 次状态转移发生的时刻，两次状态转移的时间间隔为 $T_m^n = t_m^{n+1} - t_m^n$，第 m 个再生周期中发生的状态转移次数用 n_m 表示.

记 $\{X_t^{\{i\}}; t \geq 0\} = \{X_t | X_0 = i, t \geq 0\}$ 为初始状态为 i 的一条样本轨道，$T^{\{i\}}\{i^*\} = \min\{t \geq 0 | X_t^{\{i\}} = i^*\}$ 为从初始状态 i 出发到达 i^* 的首达时间. 性能势基于样本轨道的表示式为

$$\begin{cases} g_i(\boldsymbol{\theta}) = \mathrm{E}\Big[\displaystyle\int_0^{T^{\{i\}}\{i^*\}} \big(f(X_t^{\{i\}}) - \eta(\boldsymbol{\theta})\big)\mathrm{d}t\Big] \\ g_{i^*}(\boldsymbol{\theta}) = 0 \end{cases} \tag{5}$$

对于固定的 $\boldsymbol{\theta}$，$\{X_t, t \geq 0\}$ 在每一再生周期中独立同分布，可以用下式来估计 $g_{x_{t_m^n}}(\boldsymbol{\theta})$

$$\begin{cases} \hat{g}_{x_{t_m^n}}(\boldsymbol{\theta}) = \displaystyle\sum_{k=n}^{n_m} \big(f_{x_{t_m^k}} - \hat{\eta}(\boldsymbol{\theta})\big) \cdot T_m^k \\ \hat{\eta}(\boldsymbol{\theta}) = \dfrac{1}{T_m} \displaystyle\sum_{n=1}^{n_m} f_{x_{t_m^n}} \cdot T_m^n \end{cases} \tag{6}$$

由 (4) 式和 (6) 式，通过样本轨道的第 m 个再

生周期可以得到 $\nabla\eta(\boldsymbol{\theta})$ 的一个估计

$$\widehat{\nabla\eta}_m(\boldsymbol{\theta})=\frac{1}{T_m}\sum_{n=1}^{n_m}\Big(\hat{g}_{x_{t_m^{n+1}}}(\boldsymbol{\theta})-\hat{g}_{x_{t_m^n}}(\boldsymbol{\theta})\Big)\cdot \nabla a_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})/a_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})=\frac{1}{T_m}\sum_{n=1}^{n_m}\Big(\hat{\eta}(\boldsymbol{\theta})-f_{x_{t_m^n}}\Big)\cdot T_m^n\cdot \nabla a_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})/a_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta}) \tag{7}$$

其中 $a_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})>0$，是因为若 $a_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})=0$，则 $x_{t_m^{n+1}}$ 不会出现在样本轨道中.

定理 1.　$\widehat{\nabla\eta}_m(\boldsymbol{\theta})$ 是 $\nabla\eta(\boldsymbol{\theta})$ 的一个具有有界误差的无偏估计

$$\mathrm{E}\big[\widehat{\nabla\eta}_m(\boldsymbol{\theta})\big]=\nabla\eta(\boldsymbol{\theta})$$
$$\varepsilon=\|\widehat{\nabla\eta}_m(\boldsymbol{\theta})-\nabla\eta(\boldsymbol{\theta})\|<\infty$$

证明. 由 (1) 式，$a_{ij}(\boldsymbol{\theta})$, $i,j\in S$ 有界、一阶导数有界及二次可导，则由 (3) 和 (5) 式，有 $\|\nabla\eta(\boldsymbol{\theta})\|\le c_1<\infty$，由 (7) 式，$\|\widehat{\nabla\eta}_m(\boldsymbol{\theta})\|\le c_2<\infty$，故

$$\varepsilon=\|\widehat{\nabla\eta}_m(\boldsymbol{\theta})-\nabla\eta(\boldsymbol{\theta})\|\le c_1+c_2<\infty$$

记 $\boldsymbol{l}_{ij}(\boldsymbol{\theta})=\nabla a_{ij}(\boldsymbol{\theta})/a_{ij}(\boldsymbol{\theta})$，$\mathcal{F}_\tau=\sigma\{X_t,\,0\le t\le\tau\}$，有

$$\mathrm{E}\left[\hat{\eta}(\boldsymbol{\theta})|\mathcal{F}_{t_m}\right]=\mathrm{E}\Big[\frac{1}{T_m}\sum_{n=1}^{n_m}f_{x_{t_m^n}}\cdot T_m^n\Big|\mathcal{F}_{t_m}\Big]=\eta(\boldsymbol{\theta})$$

$$\mathrm{E}\left[\hat{g}_{x_{t_m^n}}(\boldsymbol{\theta})|\mathcal{F}_{t_m^n}\right]=\mathrm{E}\Big[\sum_{k=n}^{n_m}\Big(f_{x_{t_m^k}}-\hat{\eta}(\boldsymbol{\theta})\Big)\cdot T_m^k\Big|\mathcal{F}_{t_m^n}\Big]=g_{x_{t_m^n}}(\boldsymbol{\theta})$$

$$\mathrm{E}\big[\widehat{\nabla\eta}_m(\boldsymbol{\theta})\big]=\mathrm{E}\Big[\frac{1}{T_m}\sum_{n=1}^{n_m}\Big(\hat{\eta}(\boldsymbol{\theta})-f_{x_{t_m^n}}\Big)\cdot T_m^n\cdot\boldsymbol{l}_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})\Big|\mathcal{F}_{t_m}\Big]=\mathrm{E}\Big[\frac{1}{T_m}\sum_{n=1}^{n_m}\Big(\sum_{k=n+1}^{n_m}\big(f_{x_{t_m^k}}-\hat{\eta}(\boldsymbol{\theta})\big)\cdot T_m^k-\sum_{k=n}^{n_m}\big(f_{x_{t_m^k}}-\hat{\eta}(\boldsymbol{\theta})\big)\cdot T_m^k\Big)\cdot\boldsymbol{l}_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})\Big|\mathcal{F}_{t_m}\Big]=$$

$$\mathrm{E}\Big[\frac{1}{T_m}\sum_{n=1}^{n_m}\Big(E\big[\hat{g}_{x_{t_m^{n+1}}}(\boldsymbol{\theta})\big|\mathcal{F}_{t_m^{n+1}}\big]-\mathrm{E}\big[\hat{g}_{x_{t_m^n}}(\boldsymbol{\theta})\big|\mathcal{F}_{t_m^n}\big]\Big)\cdot\boldsymbol{l}_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})\Big|\mathcal{F}_{t_m}\Big]=\mathrm{E}\Big[\sum_{n=1}^{n_m}\big(g_{x_{t_m^{n+1}}}(\boldsymbol{\theta})-g_{x_{t_m^n}}(\boldsymbol{\theta})\big)\cdot\boldsymbol{l}_{x_{t_m^n}x_{t_m^{n+1}}}(\boldsymbol{\theta})\Big|\mathcal{F}_{t_m}\Big]=\sum_{i\in S}p_i(\boldsymbol{\theta})\sum_{j\in S_i}a_{ij}(\boldsymbol{\theta})\cdot\boldsymbol{l}_{ij}(\boldsymbol{\theta})\cdot(g_j(\boldsymbol{\theta})-g_i(\boldsymbol{\theta}))=\sum_{i\in S}p_i(\boldsymbol{\theta})\sum_{j\in S}\nabla a_{ij}(\boldsymbol{\theta})\cdot g_j(\boldsymbol{\theta})=\nabla\eta(\boldsymbol{\theta})$$　□

3.2　随机逼近优化策略

动态电源管理问题的随机切换模型具有参数化的策略 $\boldsymbol{\theta}=(\theta_{ss'}^{rq},r\in S_R,s,s'\in S_S,q\in S_Q)^{\mathrm{T}}$，性能测度 $\eta_p(\boldsymbol{\theta})$ 与功耗测度 $\eta_c(\boldsymbol{\theta})$ 为参数 $\boldsymbol{\theta}$ 的连续可微函数，由 (1) 和 (3) 式，有

$$\frac{\partial\eta(\boldsymbol{\theta})}{\partial\theta_{ss'}^{rq'}}=\sum_{rsq\in S}p_{rsq}(\boldsymbol{\theta})\sum_{r's'q'\in S}\frac{\partial a_{(rsq)(r's'q')}(\boldsymbol{\theta})}{\partial\theta_{ss'}^{rq'}}g_{r's'q'}(\boldsymbol{\theta})=\sum_{q\in S_Q,q\ne q'}p_{rsq}(\boldsymbol{\theta})\cdot a_{qq'}^{rs}\cdot g_{rs'q'}(\boldsymbol{\theta})$$

式中 $p_{rsq}(\boldsymbol{\theta})\cdot a_{qq'}^{rs}\ge 0$, $q\ne q'$，且不全为 0. 故当且仅当 $g_{rs'q'}(\boldsymbol{\theta})=0$，该梯度分量为 0. 而 $g_{rs'q'}(\boldsymbol{\theta})$, $rs'q'\in S$ 当性能函数 f_p (功耗函数 f_c) 不为恒值函数时，不全为 0. 即 $\forall\,\boldsymbol{\theta}\in\Theta,\nabla\eta(\boldsymbol{\theta})\ne 0$. 故采用策略梯度法进行策略优化能够达到全局最优.

基于系统实际运行这一样本轨道的一个再生周期，通过在线学习，得到性能梯度的一个带随机误差的无偏估计 $\widehat{\nabla\eta}_m(\boldsymbol{\theta})$，结合如下带约束的 Robbins-Monro 随机逼近算法[15]，即可在下一个再生周期的开始时刻改进策略，在线进行优化.

$$\boldsymbol{\theta}_{m+1}=\Pi_\Theta\big[\boldsymbol{\theta}_m+\gamma_m\cdot\widehat{\nabla\eta}_m(\boldsymbol{\theta}_m)\big] \tag{8}$$

式中 $\Pi_\Theta[\cdot]$ 表示到 Θ 上的投影，即将 $\boldsymbol{\theta}_{m+1}$ 约束在 Θ 中取值.

步长序列选取 $\gamma_m=1/m$, $m\ge 1$，使得

$$\sum_{m=1}^{\infty}\gamma_m=\infty,\ \sum_{m=1}^{\infty}\gamma_m^2<\infty$$

梯度估计值由下式给定：

$$\widehat{\nabla\eta}_m(\boldsymbol{\theta}_m)=\begin{cases}\widehat{\nabla\eta}_p(\boldsymbol{\theta}_m), & \boldsymbol{\theta}_m\in\Theta^{G-}\\ \widehat{\nabla\eta}_\triangle(\boldsymbol{\theta}_m), & \boldsymbol{\theta}_m\in\Theta^{\triangle}\\ -\widehat{\nabla\eta}_c(\boldsymbol{\theta}_m), & \boldsymbol{\theta}_m\in\Theta^{G+}\end{cases}$$

其中 $\Theta^{G-}=\{\boldsymbol{\theta}\mid\eta_p(\boldsymbol{\theta})\le G,\ \boldsymbol{\theta}\in\Theta\}$, $\Theta^{\triangle}=\{\boldsymbol{\theta}\mid G<\eta_p(\boldsymbol{\theta})\le G+\triangle,\ \boldsymbol{\theta}\in\Theta\}$, $\Theta^{G+}=\{\boldsymbol{\theta}\mid\eta_p(\boldsymbol{\theta})>G+\triangle,\ \boldsymbol{\theta}\in\Theta\}$, $\triangle$ 为相对于 G 较小的正数. $\widehat{\nabla\eta}_p(\boldsymbol{\theta}_m)$、$\widehat{\nabla\eta}_c(\boldsymbol{\theta}_m)$ 分别为在 $\boldsymbol{\theta}_m$ 处的性能测度和功耗测度的幺模化梯度估计值. $\widehat{\nabla\eta}_{\triangle}(\boldsymbol{\theta}_m)=\widehat{\nabla\eta}_p(\boldsymbol{\theta}_m)\big(\widehat{\nabla\eta}_c(\boldsymbol{\theta}_m)\big)^{\mathrm{T}}\widehat{\nabla\eta}_p(\boldsymbol{\theta}_m)-\widehat{\nabla\eta}_c(\boldsymbol{\theta}_m)$.

设 $\boldsymbol{\theta}^*$ 是 $\eta_c(\boldsymbol{\theta})$ 在区域 $\Theta^{G+}\bigcup\Theta^{\triangle}$ 中的极值点，可以证明[15]，对于 $\forall\,\boldsymbol{\theta}_0\in\Theta$, 由 (8) 式生成的序列 $\boldsymbol{\theta}_m\to\boldsymbol{\theta}^*$, $m\to\infty$, w.p. 1，即以概率 1 收敛于 PO 的最优策略.

切换模型具有特殊的动态结构特性，其转移速率矩阵中的元素 $a_{(rsq)(r's'q')}(\boldsymbol{\theta})=a^{rs}_{qq'}\cdot\theta^{rq'}_{ss'}, r=r', q\ne q'$，其他情况 $a_{(rsq)(r's'q')}$ 不含参数 $\boldsymbol{\theta}$. 故

$$\frac{\partial a_{(rsq)(r's'q')}(\boldsymbol{\theta})/\partial\theta^{rq'}_{ss'}}{a_{(rsq)(r's'q')}(\boldsymbol{\theta})}=\begin{cases}1/\theta^{rq'}_{ss'}, & r=r', q\ne q'\\ 0, & \text{其他}\end{cases}$$

$$a_{(rsq)(r's'q')}(\boldsymbol{\theta})>0$$

结合 (7) 式可知，该优化算法不依赖于系统的具体参数值，如到达率、服务率等，具有较强的自适应性；也无需计算各状态的性能势或其他相关量 (如 Q -因子)，有效减小计算量和所需计算存储空间，提高了算法的实时性.

4 应用仿真

笔记本电脑中硬盘的运行状态切换是动态电源管理的一个典型应用. 笔记本硬盘有读写、空闲和睡眠等多种运行状态. 硬盘进行读写服务时主轴电机高速运转和读写磁头的工作使得功耗较大. 处于空闲状态时，主轴电机一直高速旋转，功耗也较大，若有读写任务到达可以及时响应，这时可以选择将硬盘切换到睡眠状态. 睡眠状态主轴电机停止旋转，功耗很小，若有读写任务到达，则需要唤醒，切换到工作状态，才能提供服务. 运行状态间的切换需要一定的时间和电能消耗. 对硬盘进行动态电源管理，在满足设计性能要求的同时，可以有效降低系统的功耗.

表 1 仿真参数
Table 1 Simulation parameters

运行状态	功耗 (W)	转换时间 (s)
读写	2.5	NA
空闲	1.6	NA
睡眠	0.1	NA
读写 ↔ 空闲	0	0
读写 ↔ 睡眠	1.2	1.0
空闲 ↔ 睡眠	1.2	1.0

在应用中读写服务请求的到达满足 Possion 分布，到达率为 λ，对读写服务的处理时间满足指数分布，服务率为 μ，运行状态之间的转换时间近似满足指数分布，其均值为 τ_k，设该系统中队列的容量为 1. 构造该问题的随机切换模型, 状态空间 $S=\{(i,0),(a,1),(a,2),(a\to s,0),(s,0),(s\to a,1),(s\to a,2)\}$. 行动空间 $D=\{d_{is},d_{ii}\}$，其中 d_{is} 表示运行状态从空闲状态至睡眠状态的切换. 控制策略 $\boldsymbol{\theta}=(\theta^0_{is},\theta^0_{ii})$，即在转为空闲的时刻以概率 θ^0_{is} 将系统切换至睡眠状态.

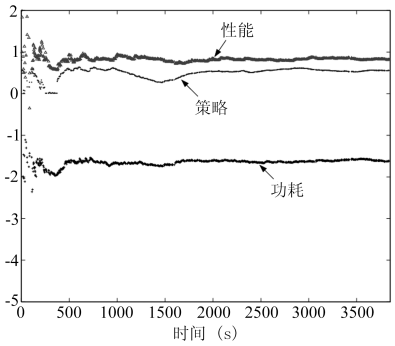

图 1 性能约束 0.82 时的在线优化结果
Fig. 1 Simulation result with performance constraint 0.82

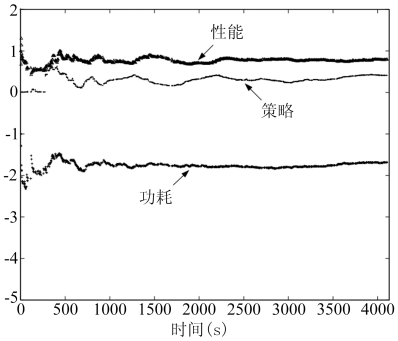

图 2 性能约束 0.80 时的在线优化结果
Fig. 2 Simulation result with performance constraint 0.80

以平均等待时间作为性能函数，优化目标是性能约束下的最小平均功耗. 运用上述在线优化算法，选取如表 1 所示仿真参数及 $\lambda=1,\mu=2$，仿真结果如图 1∼ 3 所示 (图中所示策略为 θ^0_{is}).

图 4 给出仿真结果与理论值的比较.

从仿真结果来看，采用的在线优化算法具有较快的收敛速度，结果与理论计算值相符，有效地实现了在系统参数未知情况下的策略优化.

5 结论

上述动态电源管理的随机切换模型及其在线优化算法，能够有效解决参数服从 (或近似服从) 指数分布的动态电源管理问题. 模型描述的准确性保证了优化的效果，在线优化算法可应用于系统参数难

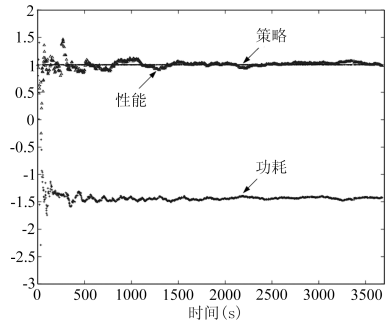

图 3 无性能约束时在线优化结果

Fig. 3 Simulation result without performance constraint

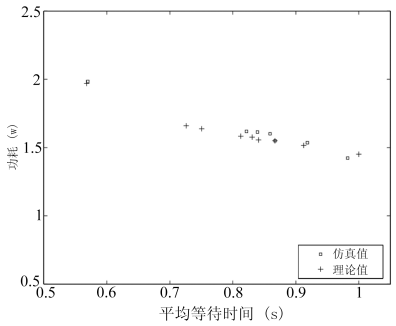

图 4 仿真结果与理论值的比较

Fig. 4 The comparison of simulation results and theoretic results

以预先获取的情况，具有运算量小，收敛速度快，较小方差，可靠精度的特点，具有较强的自适应性和较高的应用价值.

References

1 Greenawalt P M. Modeling power management for hard disks. In: Proceedings of the Second International Workshop of Modeling, Analysis, and Simulation for Computer and Telecommunication Systems. IEEE, 1994. 62∼65

2 Srivastava M B, Chandrakasan A P, Brodersen R W. Predictive system shutdown and other architectural techniques for energy efficient programmable computation. *IEEE Transaction on Very Largy Scale Integration Systems*, 1996, **4**(1): 42∼55

3 Benini L, Bogliolo A, Paleologo G A, De Micheli G. Policy optimization for dynamic power management. *IEEE Transcation on Computer Aided Design of Integrated Circuits and Systems*, 1999, **18**(6): 813∼833

4 Chung Eui-Young, Benini L, Bogliolo A, Lu Yung-Hsiang, De Micheli G. Dynamic power management for nonstationary service requests. *IEEE Transactions on Computers*, 2002, **51**(11): 1345∼1361

5 Qiu Q, Wu Q, Pedram M. Stochastic modeling of a power-managed system — construction and optimization. *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems*, 2001, **20**(10): 1200∼1217

6 Cao X R. Basic ideas for event-based optimization of Markov systems. *Discrete Event Dynamic Systems*, 2005, **15**(2): 169∼197

7 Bertsekas D P. *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 2001. **2**: 194∼200

8 Puterman M L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley, 1994. 337∼341

9 Cao X R. The potential structure of sample paths and performance sensitivities of Markov systems. *IEEE Transactions on Automatic Control*, 2004, **49**(12): 2129∼2142

10 Fang H T, Cao X R. Potential-based on-line policy iteration algorithms for Markov decision processes. *IEEE Transactions on Automatic Control*, 2004, **49**(4): 493∼505

11 Marbach P, Tsitsiklis J N. Simulation-based optimization of Markov reward processes. *IEEE Transactions on Automatic Control*, 2001, **46**(2): 191∼209

12 Chong E K P, Ramadge P J. Stochastic optimization of regenerative systems using infinitesimal perturbation analysis. *IEEE Transactions on Automatic Control*, 1994,**39**(7): 1400∼1410

13 Baxter J, Bartlett P L. Direct gradient-based reinforcement learning. In: Proceedings of IEEE International Symposium on Circuits and Systems. IEEE, 2000. 271∼274

14 Cao X R, Chen H F. Perturbation realization, potentials and sensitivity analysis of Markov processes. *IEEE Transactions on Automatic Control*, 1997, **42**(10): 1382∼1393

15 Kushner H J, Yin G. *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer, 2003. 125∼133

江 琦　中国科学技术大学自动化系博士研究生，主要研究方向为信息网络性能优化. 本文通信作者. E-mail: jiangqi@mail.ustc.edu.cn

(**JIANG Qi** Ph.D. Candidate in Department of Automation, University of Science and Technology of China. His research interests include information networks performance optimization, *etc*. Corresponding author of this paper.)

奚宏生　中国科学技术大学自动化系教授、博士生导师，主要研究领域为离散事件动态系统、信息网络性能分析与优化. E-mail: xihs@ustc.edu.cn

(**XI Hong-Sheng** Professor in Department of Automation, University of Science and Technology of China. His research interests include discrete event dynamic systems and information networks performance analysis and optimization.)

殷保群　中国科学技术大学自动化系教授、博士生导师，主要研究领域为 Markov 决策过程、离散事件动态系统优化及应用. E-mail: bqyin@ustc.edu.cn

(**YIN Bao-Qun** Professor in Department of Automation, University of Science and Technology of China. His research interests include Markov decision processes and discrete event dynamic systems performance optimization and application.)