第36卷第2期 中国科学技术大学学报 Vol. 36, No. 2
2006年2月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Feb. 2006

文章编号:0253-2778(2006)02-0183-04

# 半 Markov 控制过程基于性能势仿真的并行优化算法*

代桂平,殷保群,李衍杰,奚宏生

(中国科学技术大学自动化系,安徽合肥 230027)

**摘要**:根据等价 Markov 过程方法,研究了一类半 Markov 控制过程在紧致行动集上关于无限水平平均代价准则的性能优化算法.由于实际系统的状态空间往往非常大,因此通常的串行仿真算法可能会耗时过长,或由于硬件限制而无法实现.针对这些问题,提出了一种基于性能势的并行仿真优化算法,以期寻找系统的最优平稳策略,并用该算法对性能势的仿真和策略寻优分别进行了并行化,获得了较好的运行效率.仿真实例表明了该算法的有效性.这一算法可应用于大规模实际半 Markov 系统的性能优化.

**关键词**:半 Markov 控制过程;紧致行动集;性能势;并行仿真算法

**中图分类号**:TP202 **文献标识码**:A

# Parallel optimization algorithms for semi-markov control processes based on performance potentials simulation

DAI Gui-ping, YIN Bao-qun, LI Yan-jie, XI Hong-sheng

(*Department of Automation, USTC, Hefei* 230027, *China*)

**Abstract**: Based on the equivalent Markov process, performance optimization algorithms were studied for a class of semi-Markov control processes (SMCPs) with infinite horizon average-cost criteria and compact action set. Since the state space of a practical system is often very large, when applying traditional serial simulation algorithms, a long time is possibly required, and it is impossible to realize the algorithm due to limitations of the hardware. A parallel simulation optimization algorithm based on performance potentials was proposed to find the optimal stationary policy of a system. In this algorithm, the simulation of the performance potentials and the part of policy iteration are paralleled respectively, and high efficiency was achieved. A simulation example shows that the algorithm can get high speedup. The algorithm can be used in optimization for large-scale practical semi-Markov systems.

**Key words**: semi-Markov control processes; compact action set; performance potentials; parallel simulation algorithm

## 0 引言

半 Markov 控制过程(SMCP)是研究离散事件随机动态系统性能优化问题的一个重要模型,并在许多实际工程问题中有着广泛的应用.半 Markov 控制过程是一类受到一系列控制决策驱动的半

* **收稿日期**:2004-03-02;**修回日期**:2004-08-27
**基金项目**:国家自然科学基金(60274012, 60574065)和安徽省自然科学基金(050420301)资助.
**作者简介**:代桂平,女,1977年生,博士生.研究方向:离散事件动态系统性能灵敏度估计与优化算法. E-mail:daigping@ustc.edu
**通讯作者**:殷保群,博士/副教授. E-mail:bqyin@ustc.edu.cn

Markov 系统,其状态转移规律和控制决策所采用的行动方案相互作用决定了系统的演化,过程在每个状态的逗留时间是一个服从一般分布的随机变量.本文研究了一类半 Markov 控制过程在紧致行动集上的最优平稳策略的求解算法,从实际应用的角度出发,提出了一种基于性能势的并行仿真优化算法来寻找系统的最优平稳策略.通过仿真得到性能势的无偏估计,用于最优平稳策略的迭代寻优,获得了较好的运行效率.

## 1 问题描述

考虑一个半 Markov 过程 $Y=\{Y_t;t\geqslant 0\}$,它有一个有限的状态空间 $\Phi=\{1,2,\cdots,K\}$ 和一个行动空间 $D$,$D(i)\subset D$ 是状态 $i$ 的容许行动集.设每一个 $i\in\Phi$,$D(i)$ 为非空紧集.我们仅考虑平稳策略,一个平稳策略是一个映射 $v:\Phi\to D$,且对任意的 $i\in\Phi$,$v(i)=d_i\in D(i)$,记 $v=(v(1),v(2),\cdots,v(K))$,并令 $\Omega_s$ 是全体平稳策略集.在策略 $v$ 下,$Y$ 的半 Markov 核是 $Q^v(t)=[Q(i,j,v(i),t)]$.设在任意策略 $v\in\Omega_s$ 下,$Y$ 是不可约和非周期的,因而也是正常返的.根据文献[1]中的定理 10.5.22 可知,$Y$ 存在惟一的稳态分布 $\boldsymbol{p}^v=(p^v(1),p^v(2),\cdots,p^v(K))>0$.半过程 $Y$ 的嵌入 Markov 链 $X$ 也存在惟一的稳态分布 $\boldsymbol{\pi}^v=(\pi^v(1),\pi^v(2),\cdots,\pi^v(K))>0$,且满足 $\boldsymbol{\pi}^v\boldsymbol{e}=1$,$\boldsymbol{\pi}^v\boldsymbol{p}^v=\boldsymbol{\pi}^v$.这里,$\boldsymbol{p}^v=[p(i,j,v(i))]$ 为 $X$ 在 $v$ 下的转移矩阵,$\boldsymbol{e}=(1,1,\cdots,1)^\tau$,"$\tau$"表示转置.令 $f$ 为依赖于 $v$ 的性能函数,且对每个 $i\in\Phi$,$f(i,\cdot):D(i)\to(-\infty,+\infty)$,记 $\boldsymbol{f}^v=(f(1,v(1)),\cdots,f(K,v(K)))^\tau$.

我们称 $(Y,\Phi,D,Q^v(t),\boldsymbol{f}^v)$ 为约束在平稳策略集 $\Omega_s$ 上的一个 SMCP.$Y$ 关于无限时间水平平均代价性能准则为

$$\eta^v=\lim_{T\to\infty}\frac{1}{T}E\left\{\int_0^T f(Y_t,v(Y_t))\mathrm{d}t\right\},v\in\Omega_s \quad (1)$$

因为 $Y$ 是遍历的,故有 $\eta^v=\boldsymbol{p}^v\boldsymbol{f}^v$.优化的目标是选择一个策略 $v^*\in\Omega_s$,使得 $\eta^v$ 达到最小.

记 $Q(i,v(i),t)=\sum_{j\in\Phi}Q(i,j,v(i),t)$,

$S(i,v(i))$ 为在策略 $v$ 下过程在状态 $i$ 的平均逗留时间,即

$$S(i,v(i))=\int_0^\infty t\mathrm{d}Q(i,v(i),t) \quad (2)$$

令 $\boldsymbol{\Lambda}^v=\mathrm{diag}(S^{-1}(1,v(1)),S^{-1}(2,v(2)),\cdots,S^{-1}(K,v(K)))$,定义

$$\boldsymbol{A}^v=\boldsymbol{\Lambda}^v(\boldsymbol{p}^v-I) \quad (3)$$

则显然 $\boldsymbol{A}^v\boldsymbol{e}=0$,并可验证 $\boldsymbol{p}^v\boldsymbol{A}^v=0$.因此,$\boldsymbol{A}^v=\boldsymbol{\Lambda}^v(\boldsymbol{p}^v-I)$ 可以作为一个 Markov 过程 $\overline{X}$ 的无穷小矩阵,且 $\overline{X}$ 具有惟一的稳态分布,由上面的讨论可知,这个稳态分布就是原半 Markov 过程 $Y$ 的稳态分布,故对相同的性能函数,Markov 过程 $\overline{X}$ 和半 Markov 过程 $Y$ 在无限水平平均代价性能准则下是等价的,因此,我们可将半 Markov 过程 $Y$ 在平均代价准则下的优化问题转化为与之等价的 Markov 过程 $\overline{X}$ 的优化问题.

我们有下列假设:

**假设 1.1** 对任意 $i,j\in\Phi$,及 $t\geqslant 0$,$Q(i,j,v(i),t)$ 在 $D(i)$ 上连续.

**假设 1.2** 对任意 $i\in\Phi$,$f(i,v(i))$ 在 $D(i)$ 上连续.

对任意的 $v\in\Omega_s$,我们定义 SMCP 的平均代价 Poisson 方程为

$$(-\boldsymbol{A}^v+\boldsymbol{e}\boldsymbol{p}^v)\boldsymbol{g}^v=\boldsymbol{f}^v \quad (4)$$

这里 $\boldsymbol{p}^v$ 是 $Y$ 在策略 $v$ 下的稳态分布,也是方程

$$\boldsymbol{p}^v\boldsymbol{A}^v=0,\ \boldsymbol{p}^v\boldsymbol{e}=1 \quad (5)$$

的惟一解.易证明,矩阵 $(-\boldsymbol{A}^v+\boldsymbol{e}\boldsymbol{p}^v)$ 可逆[2],故方程(4)存在惟一解

$$\boldsymbol{g}^v=(-\boldsymbol{A}^v+\boldsymbol{e}\boldsymbol{p}^v)^{-1}\boldsymbol{f}^v \quad (6)$$

根据文献[2],我们有下列定理.

**定理 1.3** $v^*\in\Omega_s$ 是半 Markov 控制过程 $(Y,\Phi,D,Q^v(t),\boldsymbol{f}^v)$ 平均代价最优平稳策略的充分必要条件为:对任意的 $v\in\Omega_s$,有

$$\boldsymbol{f}^{v^*}+\boldsymbol{A}^{v^*}\boldsymbol{g}^{v^*}\leqslant\boldsymbol{f}^v+\boldsymbol{A}^v\boldsymbol{g}^{v^*} \quad (7)$$

式(4)可写为 $\boldsymbol{e}\boldsymbol{p}^v\boldsymbol{g}^v=\boldsymbol{f}^v+\boldsymbol{A}^v\boldsymbol{g}^v$,两边同乘 $\boldsymbol{p}^v$,且由式(5),可得 $\boldsymbol{p}^v\boldsymbol{g}^v=\boldsymbol{p}^v\boldsymbol{f}^v=\eta^v$.故我们有

$$\boldsymbol{e}\eta^v=\boldsymbol{f}^v+\boldsymbol{A}^v\boldsymbol{g}^v \quad (8)$$

于是,定理 1.3 可等价地表示为定理 1.4.

**定理 1.4** $v^*\in\Omega_s$ 是半 Markov 控制过程 $(Y,\Phi,D,Q^v(t),f^v)$ 平均代价最优平稳策略的充分必要条件是 $v^*$ 满足方程

$$0=\min_{v\in\Omega_s}\{\boldsymbol{f}^v+\boldsymbol{A}^v g^{v^*}-\boldsymbol{e}\eta^{v^*}\} \quad (9)$$

其中,$\boldsymbol{g}^v=[\boldsymbol{A}^v+\boldsymbol{e}\boldsymbol{p}^v]^{-1}\boldsymbol{f}^v=[-\Lambda^v(\boldsymbol{p}^v-I)+\boldsymbol{e}\boldsymbol{p}^v]^{-1}\boldsymbol{f}^v$ 为 Markov 过程 $\overline{X}$ 的性能势.

上式就是 SMCP 在平均代价准则下,基于等价 Markov 过程性能势的最优性方程.

由定理 1.3 可直接得到如下的策略迭代算法[3].

步 1：令 $k=0,\varepsilon>0$，选择初始策略 $v_k$；

步 2：根据式(3)计算 $\boldsymbol{A}^{v_k}$，根据式(5)和式(6)计算 $\boldsymbol{p}^{v_k}$ 和 $\boldsymbol{g}^{v_k}$；

步 3：选择 $v_{k+1}$，对每一个状态 $i\in\Phi$，满足

$$v_{k+1}(i)\in\arg\min_{v(i)\in D(i)}\Big\{f(i,v(i))+\sum_{j\in\Phi}a_{ij}(v(i))g^{v_k}(j)\Big\}\tag{10}$$

步 4：如果 $sp(\boldsymbol{f}^{v_{k+1}}+\boldsymbol{A}^{v_{k+1}}\boldsymbol{g}^{v_k})<\varepsilon$，则记 $v_\varepsilon=v_{k+1}$，算法终止；否则，置 $k:=k+1$，转步 2. 此处 $sp(h)=\max_i\{h(i)\}-\min_i\{h(i)\}$.

## 2　并行仿真算法

上面给出的策略迭代算法，每一次循环都要进行矩阵求逆运算来计算性能势，且已验证，有 90% 以上的运算量集中在这部分矩阵运算和步 3 的迭代运算中. 对于实际 SMCP 系统的优化，其状态空间往往非常巨大，上面给出的策略迭代算法中性能势的计算开销将非常大，甚至受到硬件条件限制而不可实现. 因此需要考虑发展并行算法. 我们可以对系统的一条仿真样本轨道进行分析，得到性能势的精确估计值.

$$\boldsymbol{g}^v=\hat{\boldsymbol{D}}^v\hat{\boldsymbol{\pi}}^v\tag{11}$$

其中，$\hat{D}^v=[\hat{d}_{ij}^v]$，而

$$\hat{d}_{ij}^v=\hat{E}\Big\{\int_0^{S^{\{j\}}(i)}f(\overline{X}_t^{\{j\}},v(\overline{X}_t^{\{j\}}))\mathrm{d}t\Big\}-\hat{E}\{S^{\{j\}}(i)\}\hat{\eta}^v\tag{12}$$

这里，$\overline{X}_t^{\{j\}}$ 表示初始状态为 $j$ 的 Markov 过程，$S^{\{j\}}(i)$是从状态 $j$ 到状态 $i$ 的首次到达时间. 式中，$(\hat{*})$代表$(*)$的估计值.

性能势估计的计算量在优化算法中占相当大的部分，由于状态空间较大耗时很多，因此需要对性能势的估计算法并行化，并采用螺旋式划分和公共随机数策略(CRN)以提高其并行效率[4].

另外的计算量主要集中在迭代计算. 而策略向量的迭代是分别对各个分量进行寻优，因此可以将这一部分直接并行化，即各个并行处理节点分别处理一部分分量.

针对目前比较常用的并行机和编程模型，该并行算法基于分布存储系统和消息传递模型以提高通用性. 具体利用 MPI(message passing interface)并行编程标准在曙光 2000 上实现[5]. 曙光 2000 是具有可扩展机群体系结构的通用超级并行机系统，通过显式的消息传递来交换数据，以共同完成同一计算任务. MPI 是一个并行编程标准，基于消息传递编程模型，并成为这种模型的代表和事实上的标准，可以几乎不加修改地移植到所有的并行计算机上，因此，目前应用非常广泛.

假设有 $N$ 个处理节点($N\leqslant K$)，每个处理节点对$[K/N]$([ ]代表取整运算)个状态寻优(最后一个处理节点所处理的状态数可能略少). 并行仿真算法如下：

步 1：令 $k=0,\varepsilon>0$，选择初始策略 $v_k$.

步 2：在当前策略下并行估计 $\hat{\pi}^{v_k}$ 和 $\hat{g}^{v_k}$. 进行广播操作，将 $\hat{g}^{v_k}$ 广播到各个处理节点. 其中广播操作为 MPI 中定义的标准函数，用来把变量的值复制到每一个处理结点.

步 3：在 $N$ 个处理节点中并行寻优：对每一状态 $i\in\Phi$，寻找

$$v_{k+1}(i)\in\arg\min_{v(i)\in D(i)}\Big\{f(i,v(i))+\sum_{j\in\Phi}a_{ij}(v(i))g^{v_k}(j)\Big\}$$

步 4：进行收集(gather)操作，把各个处理节点中相应的 $v_{k+1}$分量收集在一起，再进行广播(broadcast)操作，把整个 $v_{k+1}$向量广播到各个处理节点. 其中，收集操作为 MPI 中定义的标准函数，用来把各个处理结点中不同的变量收集到某个处理结点中.

步 5：如果 $sp(\boldsymbol{f}^{v_{k+1}}+\boldsymbol{A}^{v_{k+1}}\boldsymbol{g}^{v_k})<\varepsilon$，则算法停止，记 $v_{\min}=v_k$. 否则，令 $k:=k+1$，转步 2.

## 3　数值例子

考虑一个具有 $K=600$ 个状态的半 Markov 过程 $Y$，其状态空间为 $\Phi=\{1,2,\cdots,600\}$，容许行动集为紧集 $D(i)=[0.5,10]$，一个平稳策略为 $v=(v(1),\cdots,v(600)),v(i)\in D(i),i=1,2,\cdots,600$. 嵌入 Markov 链的一步转移概率为

$$p(i,1,v(i))=1-\boldsymbol{e}^{-\frac{v(i)}{i}},\ p(i,j,v(i))=\frac{1}{599}\boldsymbol{e}^{-\frac{v(i)}{i}},$$
$$i=1,2,\cdots,600;j=2,3,\cdots600$$

已知过程 $Y$ 处于状态 $i$ 下一次转移到状态 $j$，它在状态 $i$ 的逗留时间服从区间$[0,jv(i)]$上的均匀分布，即分布函数为

$$F(i,j,v(i),t)=\begin{cases}\dfrac{t}{jv(i)},0\leqslant t\leqslant jv(i),\\ 1,t>jv(i).\end{cases}$$

$Q(i,j,v(i),t)=P(i,j,v(i))F(i,j,v(i),t)$，性能

函数为 $f(i,v(i))=\ln[(1+i)v(i)]+\frac{\sqrt{i}}{2v(i)}$，$i=1,2,\cdots,600$. 显然假设 1 和假设 2 成立. 仿真结果如表 1 所示.

**表 1** 不同处理节点数目下的仿真结果

**Tab. 1** Simulation results in different node numbers

| 处理节点数目 | 性能相对误差/% | 所用时间/s |
|---|---|---|
| 1(即串行) | 0.96 | 852.12 |
| 3 | 0.97 | 333.63 |
| 5 | 0.97 | 204.21 |
| 10 | 0.98 | 138.46 |
| 15 | 0.99 | 130.00 |
| 23 | 0.99 | 119.44 |
| 35 | 0.99 | 98.20 |

实例表明并行算法的寻优结果基本不变，保持了相当的精确性，但仿真时间大大缩短. 在采用 5 个处理节点时所耗费时间仅为串行计算耗时的 24% 左右.

值得注意的是从理论上来说该算法应该具有线性的加速比，但是在实际仿真过程中只有当处理节点数目较小时才具有接近线性的加速比，而随着处理节点数目继续增多，加速比越来越小. 这有两个原因：①首先是由于各处理节点之间的通信消费造成的，通信消费随着处理节点数目增大而增大；②其次是由于程序中有无法并行化的部分，这一部分的时间开销并不随着处理节点数目的增多而减少. 因此在实际的并行优化过程中应选择合适的处理节点数目，并对程序中时间开销较大的部分进行并行化，以达到最大的消费比.

另外随着处理节点数目的增多并行算法得到的性能误差稍有增大. 这是由于并行仿真时的累积误差所致.

## 4 结论

本文在半 Markov 性能势的基础上，从实际应用的角度出发，利用等价 Markov 过程方法，给出了一类半 Markov 控制过程(SMCP)最优平稳控制策略的并行求解算法，并给出了具体的并行仿真实例. 从比较的结果来看，并行算法大大提高了效率，并在处理节点数目适当时有接近线性的加速比，因此可以应用到大规模实际半Markov系统的性能优化中.

**参考文献(References)**


[1] Cinlar E. Introduction to Stochastic Processes [M]. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

[2] XI H S, TANG H, YIN B Q. Optimal policies for a continuous time MCP with compact action set [J]. Acta Automatica Sinica, 2003, 29(2): 206- 211.

[3] DAI Gui-ping, YIN Bao-qun, LI Yan-jie, et al. Optimization algorithms for semi-Markov control process with average criteria[J]. Journal of University of Science and Technology of China, 2005, 35(2): 202-207.
代桂平，殷保群，李衍杰，等. 半 Markov 控制过程在平均代价准则下的迭代优化算法[J]. 中国科学技术大学学报，2005, 35(2): 202-207.

[4] ZOU Chang-chun, ZHOU Ya-ping, YIN Bao-qun, et al. Derivative estimates parallel simulation algorithm based on performance potentials for a class of CQNS [J]. Journal of University of Science and Technology of China, 1999, 29(1): 21-29.
邹长春，周亚平，殷保群，等. 基于性能势理论对闭排队网络进行梯度估计的并行仿真算法[J]. 中国科学技术大学学报，1999, 29(1): 21-29.

[5] Downing Information Industry Co. Ltd.. Dawning 2000 User Manual[Z]. 1998.
曙光信息产业有限公司. 曙光 2000 用户手册[Z]. 1998.