文章编号: 1001-0920(2001)04-0415-05

一类 Markov 决策过程自适应决策的新方法

李江红[1], 胡照文[2]
(1. 上海交通大学 电子信息学院, 上海 200030; 2. 中南大学 铁道校区, 湖南 长沙 410075)

摘 要: 提出一种计算效率高且能以任意给定精度实现决策近优的新方法。该方法的原理是根据要求的决策精度对参数集进行有限分区, 利用有偏极大似然估计器估计未知参数, 并在决策过程中根据估计参数所在的分区获得控制对 Markov 过程进行决策。

关键词: Markov 决策过程; 有偏极大似然估计器; 自适应决策; 近似最优
中图分类号: TP 217.2 文献标识码: A

New Algorithm for a Class of Adaptive Markov Decision Process

LI Jiang-hong[1], *HU Zhao-wen*[2]
(1. Electronic and Information College, Shanghai Jiaotong University, Shanghai 200030, China;
2. Changsha Railway College, Central South University, Changsha 410075, China)

Abstract: A new algorithm for adaptive Markov Decision Process(MDP) is proposed, which can achieve optimal decision with any desired accuracy. The principle of the new algorithm is partitioning the parameter set according to the desired accuracy and using the policy related to the partition that the new estimated parameter exists to make decision.

Key words: MDP (Markov Decision Process); biased maximum likelihood estimator; adaptive decision

1 引 言

Markov 决策过程(MDP)是应用广泛的一种随机决策过程。MDP 自适应决策是 MDP 的一个重要研究方向, 它将自适应控制的基本思想引入随机决策过程, 考虑当 Markov 过程状态转移概率依赖于未知参数时, 通过自适应决策使目标函数值趋于状态转移概率已知条件下目标函数的最优值。

MDP 自适应决策算法可分为间接决策法和直接决策法两大类。间接决策法[1~5]一般基于确定性等价原理, 即首先估计未知参数, 将参数估计值视为真实参数值; 然后计算相应的最优策略, 对 Markov 过程进行决策。直接决策法[6~8]则无参数估计, 直接对 Markov 过程进行决策。

状态空间和控制集均有限的 Markov 过程是较常见的一类 Markov 过程。文献[1, 4]研究了当影响过程状态转移概率的未知参数属于有限离散参数集时, 该类 Markov 过程的自适应决策。文献[5]则研究了未知参数属于紧集时该类 Markov 过程的自适应决策, 并证明其自适应决策算法能实现决策最优。由于每次完成参数估计后, 必须计算相对参数估计值 Markov 过程的最优策略, 而参数集中有无穷多个点, 因此自适应决策过程的计算量大, 算法效率低。

收稿日期: 2000-01-10; 修回日期: 2000-04-24
作者简介: 李江红(1970—), 男, 湖南耒阳人, 博士生, 从事随机决策、智能控制研究; 胡照文(1963—), 男, 湖南长沙人, 副教授, 从事随机决策、非线性控制研究。

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

本文针对文献[5]的不足,对状态空间和控制集均有限的MDP,提出一种计算效率高且能以任意给定精度实现近优的自适应决策算法。其基本思想是首先根据要求的决策精度,将参数集离散化为有限分区,并在每个分区上选择一个代表该分区的参数;然后利用有偏极大似然估计器估计未知参数,根据所得未知参数所在的分区得到代表该分区的参数;最后根据相对该参数Markov过程的最优策略进行决策。由于参数集分区有限并且决策过程中所需的策略在决策之前便可计算,因此极大地减少了决策过程的计算量。

2 Markov决策过程模型

记Markov过程的状态空间和控制集分别为X和U,且X和U均有限。不失一般性,设$X=\{1,2,\cdots,|X|\}$,其中$|X|$表示X中所有状态的个数。Markov过程的状态转移概率为$\{p(i,j,u,\theta^0)\}$,表示在控制u下Markov过程状态从i转移到j的概率。其中$\theta^0\in\Theta$是影响Markov过程的未知参数,Θ是已知紧集。当控制为u时,Markov过程状态从i转移到j所产生的立即损失记为$c(i,j,u)$。记$s=0,1,\cdots$时Markov过程的状态和控制分别为x_s和u_s。以θ为参数的MDP,在策略$\Phi:X\to U$下的无界期望平均值定义为

$$J(\Phi,\theta)=\lim_{N\to\infty}\frac{1}{N}E\left\{\sum_{s=0}^{N-1}c(x_s,x_{s+1},\Phi(x_s))\right\}\quad(1)$$

其最优值定义为$J^*(\theta)=\min_{\Phi\in\sum\Pi}\{J(\Phi,\theta)\}$,其中$\Pi$是所有策略组成的集合。称使$J(\Phi_\theta,\theta)=J^*(\theta)$的策略$\Phi_\theta$为最优策略。Markov决策过程的目标是使无界样本路径平均损失$\lim_{N\to\infty}\frac{1}{N}\sum_{s=0}^{N-1}c(x_s,x_{s+1},u_s)$最小。由于$\theta^0$未知,因此需要研究Markov过程的自适应决策。

本文假定上述Markov过程满足下列性质:

1) 对于任意给定的策略$\Phi:X\to U$和所有$(i,j)\in X\times X$,存在正整数$m_{i,j,\Phi}$和状态列$i=k_1,k_2,\cdots,k_{m_{i,j,\Phi}}=j$,当$t\in\{1,2,\cdots,m_{i,j,\Phi}\}$时,$p(k_t,k_{t+1},\Phi(k_t),\theta^0)>0$总成立;

2) 对于所有$(i,j)\in X\times X$和$u\in U$,$p(i,j,u,\theta)$对θ连续。

假定条件1)和2)具有一般性。由条件2)有如下引理:

引理1[5] 对于任意给定策略Φ,$J(\Phi,\theta)$和$J^*(\theta)$均是Θ上的连续函数。

3 近优自适应决策算法

对于第2节提出的Markov决策过程,本节给出近优自适应决策算法如下:

1) 确定要求的决策精度$\epsilon>0$,设定参数估计周期m,要求$m\geq m_{i,j,\Phi}+1$。

2) 根据给定决策精度ϵ划分参数集Θ。由引理1,对于任意$\theta'\in\Theta$,存在以θ'为中心的开集$O\subset\Theta$,当$\theta\in O$时,$\|J^*(\theta)-J^*(\theta')\|<\epsilon/2$,且对给定策略$\Phi$,$\|J(\Phi,\theta)-J(\Phi,\theta')\|<\epsilon/2$。由于$\Theta$是紧集,根据有限覆盖定理,总可将$\Theta$分割成有限个开集$\{\Theta_j,j=1,2,\cdots,q\}$,使上述两个不等式在$\Theta_j$中成立。在每个分区中取$\theta_j\in\Theta_j$,可得有限集$\{\theta_1,\theta_2,\cdots,\theta_l\}$,并计算以$\theta_j$为参数MDP的最优策略$\Phi_{\theta_j}=\Phi_j^*$。对于任意的$\theta\in\Theta_j$,成立

$$\|J^*(\theta_j)-J^*(\theta)\|<\epsilon/2\quad(2)$$

$$\|J^*(\Phi_{\theta_j},\theta_j)-J^*(\Phi_{\theta_j},\theta)\|<\epsilon/2\quad(3)$$

3) 参数估计。对所有$(i,j)\in X\times X$和$u\in U$,定义

$$n_t(i,j,u)=1+\sum_{s=0}^{t-1}I(x_s=i,x_{s+1}=j,u_s=u)\quad(4)$$

其中$I(\cdot)$为集合的特征函数。对于任意$\theta\in\Theta$,定义

$$\bar{D}_t(\theta)=\exp(J^*(\theta))^{-o(t)}\times\prod_{(i,j)\in X\times X,u\in U}p(i,j,u,\theta)^{n_t(i,j,u)}\quad(5)$$

要求$\bar{D}_t(\theta)$中$o(t)$在$\beta\in(0,1)$时,满足$0<\lim_{t\to\infty}\frac{o(t)}{t^\beta}<\infty$,在$t=0,m,2m,\cdots$时,由$\hat{\theta}_t=\arg\max_{\theta\in\Theta}\bar{D}_t(\theta)$计算未知参数估计值$\hat{\theta}_t$;在$km\leq t<(k+1)m$时,取$\hat{\theta}_t=\hat{\theta}_{km}$。这里的参数估计方法与文献[5]的方法完全相同。

4) 决策。由t时的参数估计值$\hat{\theta}_t$可得代表$\hat{\theta}_t$所在分区的参数$\alpha_t\in\{\theta_1,\theta_2,\cdots,\theta_l\}$,以及相应的最优策略$\Phi_{\alpha_t}$。令决策策略$\Phi(\cdot,\hat{\theta}_t)=\Phi_{\alpha_t}$,根据$\Phi(x_t,\hat{\theta}_t)$计算控制$u_t=\Phi(x_t,\hat{\theta}_t)=\Phi_{\alpha_t}(x_t)$对Markov过程决策。这里的决策方法与文献[5]中决策方法不同,后者需要先计算以$\hat{\theta}_t$为参数的MDP最优策略,再据此对Markov过程进行决策。

上述自适应决策算法的目的在于获得参数α_t,然后根据α_t进行决策。由于可事先确定所有的可能策略$\{\Phi_1^*,\Phi_2^*,\cdots,\Phi_l^*\}$,因而在以后的决策过程中每

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

需计算策略, 从而提高了计算效率。

4　自适应决策算法性能分析

记自适应决策策略下的概率空间为(Ω, F, P), 其中 P 是相对以 θ^0 为参数实际系统的状态转移概率, F_t 是由$(x_0, x_1, \cdots, x_t)$ 生成的 σ 代数。一般而言, $x_t(\cdot), u_t(\cdot), \hat{\theta}_t(\cdot)$ 以及它们的复合都是随机变量, 即它们都是 $\omega\in\Omega$的可测函数。在下面分析过程中, 将不显示地表示出这种关系。

定义 $I(\cdot)$ 为集合的特征函数。对于上面的自适应决策方法, 已有以下一些结论[5]:

引理 **2**　存在 $N_1\subset\Omega, P(N_1)=0$, 当 $\omega\notin N_1$ 时, $\lim\limits_{t\to\infty}\sup J^*(\hat{\theta}_t)\le J^*(\theta^0)$。

引理 **3**　存在 $N_4\subset\Omega, P(N_4)=0$, 当 $\omega\notin N_4$ 时, 如果有策略 $\Psi: X\to U$ 满足

$$\lim_{t\to\infty}\sup\frac{1}{t}\sum_{s=0}^{t-1}I(\Phi(\cdot,\hat{\theta}_s)=\Psi)>0 \qquad (6)$$

则存在$\{t_k\}$ 对所有$(i,j)\in X\times X$, $\lim\limits_{k\to\infty}p(i,j,\Psi(i),\hat{\theta}_{t_k})=p(i,j,\Psi(i),\theta^0)$ 成立, 且对每个k, $\Phi(\cdot,\hat{\theta}_{t_k})=\Psi$。

上述引理的详细证明参见文献[5]。从证明过程可以发现, 它们只与参数辨识方法有关, 而与控制的具体形式无关。由于本文的参数辨识方法与文献[5]相同, 因此上述引理对本文的自适应决策算法也成立。

根据前面的自适应算法, 由参数分区的有限性可知存在 $\Phi_{\alpha^*}\in\{\Phi_{\theta_1},\Phi_{\theta_2},\cdots,\Phi_{\theta_q}\}$, 满足

$$\lim_{s\to\infty}\sup\frac{1}{s}\sum_{t=0}^{s-1}I(\Phi(\cdot,\hat{\theta}_s)=\Phi_{\alpha^*})>0$$

由引理2, 当$\omega\notin N_4$时存在$\{t_k\}$, 对所有$(i,j)\in X\times X$, 有

$$\lim_{k\to\infty}p(i,j,\Phi_{\alpha^*}(i),\hat{\theta}_{t_k})=p(i,j,\Phi_{\alpha^*}(i),\theta^0) \quad (7)$$

定理 **1**　当 $\omega\notin N_4$ 时, $\lim\limits_{k\to\infty}\sup J(\Phi_{\alpha^*},\hat{\theta}_{t_k})=J(\Phi_{\alpha^*},\theta^0)$。

证明　由于策略Φ下MDP的无界期望平均损失 $J(\Phi,\theta)$ 可表示为[4]

$$J(\Phi,\theta)=\sum_{x\in A}\pi(x,\Phi,\theta)\sum_{y\in A}p(x,y,\Phi(x),\theta)c(x,y,\Phi(x)) \qquad (8)$$

记 $\pi(\Phi,\theta)=[\pi(1,\Phi,\theta),\cdots,\pi(|X|,\Phi,\theta)]_{|X|\times 1}$, $\pi(\Phi,\theta)$ 唯一且满足方程 $\pi(\Phi,\theta)=\pi(\Phi,\theta)P(\Phi,\theta)$, 其中 $P(\Phi,\theta)=[p(i,j,\Phi(i),\theta)]_{|X|\times|X|}$。当 $\omega\notin N_4$ 时, 由式(7) 可得$\lim\limits_{k\to\infty}P(\Phi_{\alpha^*},\hat{\theta}_{t_k})=P(\Phi_{\alpha^*},\theta^0)$。设$\pi^*$是$\{\pi(\Phi_{\alpha^*},\hat{\theta}_{t_k})\}$ 的任意极限, 则存在$\{t_k\}$ 的子列$\{t_l\}$, 满足

$$\lim_{l\to\infty}\pi(\Phi_{\alpha^*},\hat{\theta}_{t_l})[P(\Phi_{\alpha^*},\hat{\theta}_{t_l})-I]=\pi^*[P(\Phi_{\alpha^*},\theta^0)-I]=0 \qquad (9)$$

由π的唯一性得 $\pi^*=\pi(\Phi_{\alpha^*},\theta^0)$, 再由$\pi^*$ 的任意性得$\lim\limits_{k\to\infty}\pi(\Phi_{\alpha^*},\hat{\theta}_{t_k})=\pi(\Phi_{\alpha^*},\theta^0)$。故

$$\lim_{k\to\infty}\sup J(\Phi_{\alpha^*},\hat{\theta}_{t_k})=\lim_{k\to\infty}\sup\sum_{i\in A}\pi(i,\Phi_{\alpha^*},\hat{\theta}_{t_k})\times\sum_{j\in A}p(i,j,\Phi_{\alpha^*}(i),\hat{\theta}_{t_k})c(i,j,\Phi_{\alpha^*}(i))=\sum_{i\in A}\pi(i,\Phi_{\alpha^*},\theta^0)\sum_{j\in A}p(i,j,\Phi_{\alpha^*}(i),\theta^0)\times c(i,j,\Phi_{\alpha^*}(i))=J(\Phi_{\alpha^*},\theta^0)$$

(证毕)

由于 $J^*(\theta^0)$ 是参数为 θ^0 时的最优值, 因此 $J(\Phi_{\alpha^*},\theta^0)\ge J^*(\theta^0)$。

定理 **2**　记 $N=N_1\cup N_4$, 当 $\omega\notin N$ 时

$$\lim_{k\to\infty}\sup J(\Phi_{\alpha^*},\hat{\theta}_{t_k})\le J^*(\theta^0)+\epsilon$$

其中 ϵ 是给定的决策精度。

证明略。

由定理1和定理2知, 存在$0\le\xi\le\epsilon$, 使$J(\Phi_{\alpha^*},\theta^0)=J^*(\theta^0)+\xi$。

5　自适应决策误差分析

由于对 $J^*(\theta^0)$ 存在有界向量 $w(\theta^0)=[w(1,\theta^0),\cdots,w(|X|,\theta^0)]_{|X|\times 1}$, 对任意 $i\in X$ 满足[5]

$$J^*(\theta^0)=\sum_{j\in A}p(i,j,\Phi_{\theta^0}(i),\theta^0)[c(i,j,\Phi_{\theta^0}(i))+w(j,\theta^0)]-w(i,\theta^0)$$

其中 Φ_{θ^0} 是真实 MDP 的最优策略。因此定义

$$g(i,u)=\sum_{j\in A}p(i,j,u,\theta^0)[c(i,j,u)+w(j,\theta^0)]-J^*(\theta^0)-w(i,\theta^0) \qquad (10)$$

用它表示在控制 u 下真实 MDP 的目标函数值及其最优值之间的偏差。由定理 1 和定理 2 知, 当 $\Phi=\Phi_{\alpha^*}$ 时 $g(i,\Phi(i))\le\epsilon$, 因此 Φ_{α^*} 是ϵ近似最优策略。由 $g(\cdot,\cdot)$ 定义, 有

$$\Gamma(i)=\{u: g(i,u)\le\epsilon\} \qquad (11)$$

$$\Gamma=\{\Phi: \text{对所有 } i\in x, g(i,\Phi(i))\le\epsilon\} \qquad (12)$$

显然, $\Gamma(i)$ 是状态为 i 时的所有 ϵ 近似最优控制, Γ

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

是所有 ϵ 近似最优策略。

在式(10) ～(12) 的基础上, 类似于文献[5] 中定理 9 的证明可得如下结论:

定理 3 近优自适应决策算法下的控制 $\{u_t\}$ 满足

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}I(u_s\in\Gamma(x_s))=1,\quad \text{a.s.}\qquad(13)$$

定理 3 表明在自适应决策过程中, 几乎所有的控制都是 ϵ 近似最优控制。

定理 4 近优自适应决策算法下的无界样本路径平均损失

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}c(x_s,x_{s+1},u_s)\le J^*(\theta^0)+\epsilon,\quad \text{a.s.}\qquad(14)$$

证明 由 $\Gamma(x)$ 的定义可知, 对所有 $x\in X$, 当 $u\in\Gamma(x)$ 时 $g(x,u)\le\epsilon$。因此在近优自适应决策算法下, 由定理 3 有 $\lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}g(x_s,u_s)\le\epsilon$, a.s.。定义

$$f_{s+1}=c(x_s,x_{s+1},u_s)-J^*(\theta^0)-w(x_s,\theta^0)+w(x_{s+1},\theta^0)-g(x_s,u_s)$$

易证 $E[f_{s+1}|F_s]=0$, 因此 $\{f_{s+1},F_s\}$ 是鞅差列。由鞅稳定定理[9], 有 $\lim_{t\to\infty}\sup\frac{1}{t}\sum_{s=1}^{t-1}f_s=0$, a.s., 即

$$\lim_{t\to\infty}\sup\frac{1}{t}\sum_{s=0}^{t-1}c(x_s,x_{s+1},u_s)-J^*(\theta^0)-\lim_{t\to\infty}\sup\frac{1}{t}\sum_{s=0}^{t-1}g(x_s,u_s)+\lim_{t\to\infty}\sup\frac{1}{t}[w(x_s,\theta^0)-w(x_0,\theta^0)]=0,\quad\text{a.s.}$$

由 $w(x,\theta^0)$ 的有界性易得式(14) 成立。(证毕)

由于在 θ^0 已知条件下, 无界样本路径平均损失最优值为 $J^*(\theta^0)$, 因此定理 4 表明, 近优自适应决策算法下的无界样本路径平均损失实际值与理想最优值之差不大于 ϵ。

6 仿真研究

文献[5] 证明了自适应决策算法的有效性, 但未给出应用实例。下面通过仿真比较[5] 和本文提出的自适应决策算法的性能。这里称[5] 中算法为 Kumar 算法。

仿真研究的未知 MDP 如下: 状态空间 $X=\{1,2\}$, 控制集 $U=\{1,2,3,4,5\}$, 立即损失 $c(i,j,u)=12+(2-i)(7.8-0.3u-6j)$。状态转移概率 $\{p(i,j,u,\theta)$ 未知, 但有

$$p(1,2,1,\theta)=p(2,1,5,\theta)=0.01(-2.79\theta^2+4.8\theta+93.76)$$
$$p(2,2,1,\theta)=p(1,1,1,\theta)=0.01(1.98\theta^2-0.58\theta+0.60)$$
$$p(1,1,2,\theta)=p(2,2,4,\theta)=0.01(1.79\theta^2-0.31\theta+3.24)$$
$$p(2,2,2,\theta)=p(2,2,3,\theta)=0.01(1.878\theta^2-0.21\theta+2.64)$$
$$p(1,1,3,\theta)=p(1,2,5,\theta)=0.01(1.74\theta^2+0.77\theta+5.23)$$

且 $\theta\in\Theta=[-3,3]$。决策目标函数是无界样本路径平均损失。图 1 给出了无界期望平均损失最优值 $J^*(\theta)$ 在参数集 Θ 上的变化。

对该未知 MDP, 根据文献[5] 和本文算法得到的自适应决策仿真结果列于表1, 其中给定决策精度 $\epsilon=0.2$。根据该精度, 参数集 Θ 在仿真之前等分为 64 个子区间。表 1 中的仿真时间、平均损失、估计参数都是它们在决策规划水平 $N=10\,000$ 时的取值; 效率提高 = (Kumar 算法仿真时间 - 本文算法仿真时间) ÷ Kumar 算法仿真时间。注意到表 1 中有的估计参数距离 θ^0 并不很近, 这是有偏极大似然估计的特点。图 2 给出了当 $\theta^0=2$ 时两种算法下的自适应决策估计参数的变化。

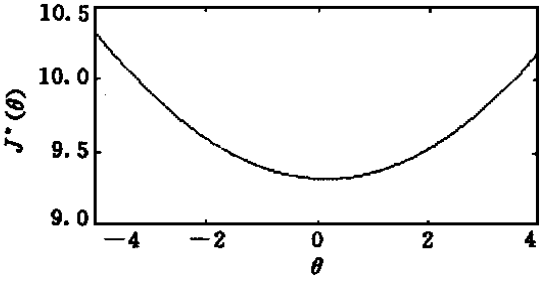

图 1 $J^*(\theta)$ 在 Θ 上的变化

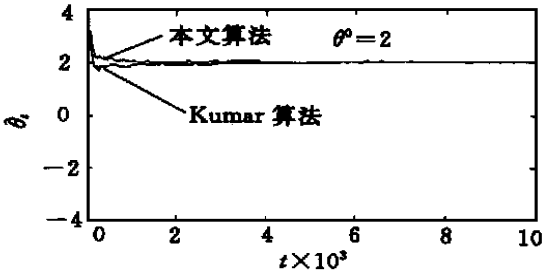

图 2 自适应决策中 $\hat{\theta}_t$ 的变化

从仿真结果不难发现, 本文提出的算法能够实现自适应近优决策。同文献[5] 的算法相比, 本文算法的计算效率有明显的提高, 其原因是在决策过程

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 1　自适应决策仿真结果

仿真参数		Kumar 算法			本文算法			效率提高
θ^0	$J^*(\theta^0)$	仿真时间(s)	平均损失	估计参数	仿真时间(s)	平均损失	估计参数	(%)
2.0	9.521	2 446	9.531	2.019 4	1 653	9.521	2.012 8	32
1.0	9.360	2 653	9.371	1.092 8	1 739	9.537	0.963 5	34
0.0	9.318	2 366	9.318	0.003 5	1 555	9.323	0.020 8	34
−1.0	9.395	2 582	9.402	1.338 1	1 747	9.395	1.209 8	32
−2.0	9.591	2 435	9.591	2.302 2	1 624	9.582	2.232 5	33

中不需要计算最优策略。由于采用策略迭代法计算最优策略,随着 MDP 状态数目和控制数目的增加,算法效率的提高将会更加显著。对于状态数和控制数均为 5 的 MDP,仿真结果表明本文算法效率提高大于 50%。

7　结　　语

在 MDP 自适应决策中,当影响状态转移概率的未知参数属于有限离散参数集时,现有方法已圆满地解决了自适应最优决策问题。但是对于紧参数集,现有方法存在计算量大,决策效率低等不足。针对这一问题,本文在现有研究成果基础上,对该类 Markov 决策过程自适应决策进行分析,提出了解决原有算法的不足并能以任意给定的精度实现决策自优的自适应决策方法。

参考文献:

[1] Borkar V, Varaiya P. Adaptive control of Markov chains (Ⅰ): Finite parameter set[J]. IEEE Trans on Autom Contr, 1979, 24(6): 953–958.

[2] Doshi B, Shreve S E. Strong consistency of a modified maximum likelihood estimator for controlled Markov chains[J]. J of Appl Probab, 1980, 17(3): 726–734.

[3] Borkar V, Varaiya P. Identification and adaptive control of Markov chains[J]. SIAM J Contr and Opt, 1982, 20(4): 470–489.

[4] Kumar P R, Becker A. A new family of optimal adaptive controllers for Markov chains[J]. IEEE Trans on Autom Contr, 1982, 27(1): 137–146.

[5] Kumar P R, Lin W. Optimal adaptive controllers for unknown Markov chains[J]. IEEE Trans on Autom Contr, 1982, 27(4): 765–774.

[6] Wheeler R M, Narendra K S. Decentralized learning in finite Markov chains[J]. IEEE Trans on Autom Contr, 1986, 31(6): 519–526.

[7] Watkins C J, Dayan P Q. Learning[J]. Machine Learning, 1992, 8(2): 279–292.

[8] Santharam G, Sastry P S. A reinforcement learning neural network for adaptive control of Markov chains[J]. IEEE Trans on Syst, Man and Cyb– Part A: Syst and Humans, 1997, 27(5): 588–600.

[9] Loeve M. Probability theory (Ⅱ)[M]. New York: Springer–Verlag, 1978.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net