

文章编号: 1001-0920(2004)09-0983-05

马尔可夫决策过程复杂性的熵测度

王红卫, 李 琛, 刘会新

(华中科技大学 系统工程研究所, 湖北 武汉 430074)

摘 要: 应用 Shannon 熵和其他熵指数来度量马尔可夫决策的复杂性. 将马尔可夫链的复杂性、不确定性和不可预测性的度量扩展到马尔可夫决策, 提出一套基于信息理论的复杂性度量方法, 可用于随机和确定性策略下的完全观测和不完全观测马尔可夫决策. 对有关数值进行仿真研究, 并给出了计算结果.

关键词: 马尔可夫决策过程; 复杂性; 熵测度

中图分类号: TP14 **文献标识码:** A

Entropic measurements of complexity for Markov decision processes

WANG Hongwei, LI Chen, LIU Hui-xin

(Institute of Systems Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Correspondent: WANG Hongwei, E-mail: wanghw@whu.edu.cn

Abstract: Shannon's entropy is applied to measure the complexity of Markov decision processes. The measurements of the complexity, uncertainty and unpredictability for Markov chains are extended to the measurements for Markov decision processes. A method of complexity measurements based on information theory is developed, and it can be used for fully and partially observable Markov decision processes under randomized and deterministic policies. A numerical result is given out to illustrate the method.

Key words: Markov decision processes; complexity; entropic measurement

1 引 言

近年来,许多学者致力于研究如何描述使系统变得复杂的复杂动态行为和混沌行为,例如美国 Santa Fe 研究所提出的复杂适应系统理论.由于复杂性的多面性,其定义^[1]还不能被多数学者所接受,而如何度量复杂性则是始终没有解决的问题.实际上,不同的复杂性度量方法是以不同的方法来度量一个复杂系统的难度,包括构造一个对象,描述一个系统,执行一个任务等.例如一种称为算法复杂性或计算复杂性,它是用来度量准确重构一个序列的难度^[1].另一种复杂性的直觉概念是不确定性、随机

性和不可预测性的度量.在复杂系统中,不确定性、随机性和不可预测性会加大求解问题的困难.

熵最早由 Shannon 引入信息理论,并应用于信息编码,他把熵定义为描述系统状态或系统不确定水平所需要的信息总量. Kolmogorov 将 Shannon 熵理论扩展到动态过程的研究,并揭示了信息与一种复杂性的内在联系.后来的研究者提出了一些有关熵测度的概念,如块熵、熵率、熵增加、可预测性增加、过渡熵和瞬时信息等^[1~6].这些熵测度用来度量不同序列的复杂性、不确定性和不可预测性,并在时间序列、排队过程、字母序列和马尔可夫链等得到应

收稿日期: 2003-09-08; 修回日期: 2003-11-19

基金项目: 教育部优秀青年教师教学和科研奖励基金资助项目

作者简介: 王红卫(1965—),男,浙江宁波人,教授,博士生导师,从事供应链管理、复杂系统的建模与仿真的研究;
李琛(1979—),女,湖北武汉人,博士生,从事供应链动态智能决策与分析的研究

用

如何度量复杂系统中动态决策的复杂性是一个非常... 熵率、熵增加和可预测性增加等熵测度方法...

2 随机过程的熵测度

本文考虑的随机过程是时不变随机变量序列... 下面讨论随机过程的熵测度[7].

1) 块熵: 长度为L 的序列的块熵定义为

H(L) = H(s1, s2, ..., sL) = H(s1) + H(s2 | s1) + H(s3 | s2, s1) + ... + H(sL | sL-1, ..., s1) = - sum_{s^L} Pr(s^L) log2 Pr(s^L).

其中: L > 0; 当 L = 0 时, 定义 H(0) = 0 它可用来度量随机序列的不确定性

2) 熵增加和熵率: 熵增加 h_mu(L) 定义为块熵 H(L) 的一阶导数, 即

h_mu(L) = dH(L) / dL = H(L) - H(L-1), L > 0

当 L = 0 时, 定义 h_mu(0) = log2 |Sigma|, log2 |S| 是单个随机变量的最大可能熵 h_mu(L) 也可表示成信息增加或条件熵, 即

h_mu(L) = Delta H(L) = D [Pr(s^L) - Pr(s^{L-1})] = H[S_L | S^{L-1}]

其中 h_mu(L) 是序列源随机性的估计

定义另一个有限熵估计为

h_mu(L) = H(L) / L, L > 0

当 L = 0 时, 同样定义 h_mu(0) = log2 |S|

定义熵率为块熵关于 L 的增长率, 即

h_mu = lim_{L -> infinity} h_mu(L) = lim_{L -> infinity} H(L) / L

熵率是随机性的测度, 可量化序列的不可归约的随机性

3) 可预测性增加: 可预测性增加定义为块熵的二阶导数, 即

Delta^2 H(L) = Delta h_mu(L) = h_mu(L) - h_mu(L-1), L > 1

其中 Delta^2 H(L) 表示不可预测性的减少率, 而 -Delta^2 H(L) 可看作从包含 L-1 个变量的序列到包含 L 个变量的序列的可预测性的增加 总可预测性增加定义为

G = sum_{L=1} Delta^2 H(L)

3 完全观测的马尔可夫决策过程的熵测度

在马尔可夫决策过程中, 决策者在每个时间单元时刻会选择一种行动, 以此推进系统的状态变化 马尔可夫决策可分为完全观测马尔可夫决策和不完全观测马尔可夫决策

完全观测马尔可夫决策过程是马尔可夫链的简单扩展, 其马尔可夫链不是自主变化的, 而是由决策者选择的行动推进 设 A 是行动集, 并假定为有限集, a_n in A 表示时刻 n 所选择的行动

设 P^a(a in A) 表示与行动有关的转移概率矩阵 从时刻 n 到下一决策时刻 n+1, 马尔可夫链将按照与行动 a_n 相关的转移概率矩阵推进 转移概率矩阵的元素可表示为

P_ij^a = Pr(s_{n+1} = j | s_n = i, a_n = a), i, j in S

设 W^a(a in A) 表示与行动有关的报酬矩阵 报酬矩阵的元素 W_ij^a 表示在时刻 n 的状态为 s_n = i, 所选择的行动为 a, 且下一时刻的状态为 s_{n+1} = j 时所得到的报酬 假定报酬矩阵是有限的, 并设 W_i^a =

sum_{j=1}^k P_ij^a W_ij^a 为时刻 n 的状态为 s_n = i 和所选择的行动为 a 的期望报酬

3.1 随机策略

在完全观测的马尔可夫决策过程中, 假设决策者在决策时刻执行某一行动可表示为一个随机变量 这样便存在一个行动集的概率分布, 其概率与系统的状态有关, 可以表示为 R_i^a, 且 sum_{a in A} R_i^a = 1. 在时不变情况下, 转移概率矩阵变为 sum_{a in A} R_i^a P_ij^a. 假设 P_i 为状态的均衡概率, 这样便可计算它的熵测度

根据式(1) 和马尔可夫的块熵公式, 便可得到在随机策略下的完全观测的马尔可夫决策过程的块熵, 即

H(L) = H(s1, s2, ..., sL) = H(s1) + H(s2 | s1, a1) + H(s3 | s2, a2) + ... + H(sL | sL-1, aL-1)

$$\begin{aligned}
 & - \sum_{i=1}^k P_i \log_2 P_i - (L-1) \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a \\
 & \log_2 \Pr(s_1) \Pr(s_2 | s_1, a_1) \dots \Pr(s_L | s_{L-1}, a_{L-1}) = \\
 & - \sum_{i=1}^k P_i \log_2 P_i - (L-1) \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a
 \end{aligned} \tag{9}$$

根据式(9), 可得到其他熵测度

$$h_\mu(L) = \begin{cases} - \sum_{i=1}^k P_i \log_2 P_i, L = 1; \\ - \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a, L = 2; \end{cases} \tag{10}$$

$$\begin{aligned}
 h_\mu(L) = & - \frac{1}{L} \sum_{i=1}^k P_i \log_2 P_i - \\
 & \frac{(L-1)}{L} \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a, L > 0; \tag{11}
 \end{aligned}$$

$$h_\mu = - \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a; \tag{12}$$

$$\Delta^2 H(L) = \begin{cases} - \sum_{i=1}^k P_i \log_2 P_i - \log_2 |\mathbf{S}|, L = 1; \\ - \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a + \sum_{i=1}^k P_i \log_2 P_i, L = 2; \\ 0, L = 3; \end{cases} \tag{13}$$

$$G = - \sum_{i,j=1}^k P_i R_i^a P_{ij}^a \log_2 P_{ij}^a - \log_2 |\mathbf{S}| \tag{14}$$

3.2 确定性策略

在第 3.1 节, 决策者执行某一行动被看作随机行为, 但在大多实际情况下, 决策者选择某一行动是依据一个确定的策略或在不同决策时刻依据不同的确定性策略, 例如值迭代的最优策略和策略空间迭代的最优策略等^[8]。虽然一个确定性策略或确定性策略序列可能隐藏着一个行动集的概率分布, 但不知其准确的概率分布。一个确定性策略包含了决策者在每个状态时决策者所选择的行动, 它可表示成一个向量, 其元素 $e_i(n)$ 是决策时刻为 n , 所处状态为 i 的决策行动。在决策时刻 n 的策略将决定此时的转移概率, 它是整个转移概率 P_{ij}^a 中的一部分, 可表示为

$$P_{ij}(n) = P_{ij}^a | a = e_i(n). \tag{15}$$

如果在不同决策时刻采用不同的策略, 则转移概率矩阵是变化的。这些序列性的熵测度(如块熵、熵增加和可预测性增加等)的收敛性是不可知的, 相关

的指标性熵测度(如熵率和总可预测性增加等)的公式也无法得到

设 $P_i(1)$ 为初始状态的概率, 则两个相邻状态的概率之间的关系为

$$P_j(L) = \sum_{i=1}^k P_i(L-1) P_{ij}(L-1), \tag{16}$$

这样便可得到这些序列性的熵测度, 块熵为

$$\begin{aligned}
 H(L) &= H(s_1, s_2, \dots, s_L) = \\
 H(s_1) &+ H(s_2 | s_1, a_1) + H(s_3 | s_2, a_2) + \\
 &\dots + H(s_L | s_{L-1}, a_{L-1}) = \\
 & - \sum_{i=1}^k P_i(1) \log_2 P_i(1) - \\
 & \sum_{i,j=1}^k P_i(1) P_{ij}(1) \log_2 P_{ij}(1) - \dots - \\
 & \sum_{i,j=1}^k P_i(L-1) P_{ij}(L-1) \log_2 P_{ij}(L-1) = \\
 & - \sum_{i=1}^k P_i(1) \log_2 P_i(1) - \\
 & \sum_{l=1}^{L-1} \sum_{i,j=1}^k P_i(l) P_{ij}(l) \log_2 P_{ij}(l). \tag{17}
 \end{aligned}$$

熵增加和可预测性增加等也可根据式(2), (4) 和(6) 得到

在确定性策略情况下, 完全观测的马尔可夫决策过程指标性的熵测度没有定义。对于序列性的熵测度, 可用仿真的方式进行研究, 仿真可清楚地表明这些熵测度是不收敛的

4 不完全观测马尔可夫决策过程的熵测度

4.1 随机策略

在不完全观测的马尔可夫决策过程中, 系统的状态是不可知或无法观测的, 但存在另外一个可观测的随机变量 $d \in D$, 且它与状态序列直接有关。其关系可用一概率 q_{jd}^a 表示, 意指现有决策时刻所选择的行动为 a , 下一时刻状态为 j , 且得到的观测状态为 d 时的概率

不完全观测的马尔可夫决策过程中有两个随机序列, 其一是状态序列 s_1, s_2, \dots, s_L , 另一个是可观测变量序列 d_1, d_2, \dots, d_{L-1} , 它们之间的关系如图 1 所示。其状态序列的熵测度与完全观测的马尔可夫决策过程相同。但由于状态是不可观测变量, 讨论其熵测度无实际意义。可观测变量序列的每一项与状态序列的对应项有关, 例如 d_1 和 s_2 与在状态 s_1 时的行动相关, 但它实际上只与状态 s_1 以及在状态 s_1 时的行动相关, 如图 1 所示

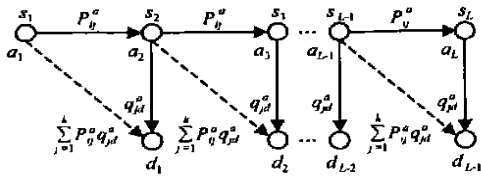


图 1 两个随机序的关系

假设不完全观测的马尔可夫决策过程是时不变的, 其状态的均衡概率为 P_i 根据式 (12), 可得到其可观测变量序列的块熵

$$\begin{aligned}
 H(L) &= H(d_1, d_2, \dots, d_L) = \\
 &H(d_1 | s_1, a_1) + H(d_2 | s_2, a_2) + \dots + H(d_L | s_L, a_L) = \\
 &-\sum_{s_1, s_2, d_1, D, a_1, A} \Pr(d_1, s_1, a_1) \log_2 \Pr(d_1 | s_1, a_1) - \\
 &\sum_{s_2, s_2, d_2, D, a_2, A} \Pr(d_2, s_2, a_2) \log_2 \Pr(d_2 | s_2, a_2) - \dots - \\
 &\sum_{s_L, s_L, d_L, D, a_L, A} \Pr(d_L, s_L, a_L) \log_2 \Pr(d_L | s_L, a_L) = \\
 &-L \sum_{d, D, i=1}^k P_i R_i^a \sum_{j=1}^k P_{ij}^a q_{jd}^a \log_2 \sum_{j=1}^k P_{ij}^a q_{jd}^a \quad (18)
 \end{aligned}$$

根据式 (2), (4) 和 (5), 通过对式 (18) 的分析, 可得到 $h_\mu(L)$, $h_\nu(L)$ 和 h_μ , 而且它们是相同的 根据式 (6), (7) 和 (18), 采用随机策略时不完全观测的马尔可夫决策过程的可预测性增加和总预测性也可得到

4.2 确定性策略

在确定性策略情况下, 仍主要研究可观测变量序列 d_1, d_2, \dots, d_{L-1} 与完全观测的马尔可夫决策过程一样, 如果在不同决策时刻采用不同策略, 则从状态到可观测变量的转移概率矩阵是变化的 从状态到可观测变量的转移概率与此时的行动、状态转移概率 P_{ij}^a 以及概率 q_{jd}^a 有关, 可表示为

$$P_{id}(n) = \sum_{j=1}^k P_{ij}^a q_{jd}^a | a = e_i(n). \quad (19)$$

在决策时刻 n , 状态转移概率仍是整个状态转移概率 P_{ij}^a 的一部分, 可由式 (15) 表示; 两个相邻状态的概率间关系可由式 (16) 确定 这样, 如果初始状态的概率已知, 便可计算任何时刻的状态概率和可观测变量的概率

于是可计算出序列性的熵测度, 但由于这些序列性的熵测度的收敛性不可知, 从而无法得到指标性的熵测度 根据第 3.1 节和第 3.2 节, 可得到块熵、熵增加和可预测性增加等熵测度, 即

$$H(L) = H(d_1, d_2, \dots, d_L) =$$

$$\begin{aligned}
 &H(d_1 | s_2, a_1) + H(d_2 | s_3, a_2) + \dots \\
 &+ H(d_L | s_{L+1}, a_L) = \\
 &H(d_1 | s_1, a_1) + H(d_2 | s_2, a_2) + \dots \\
 &+ H(d_L | s_L, a_L) = \\
 &-\sum_{s_1, s_2, d_1, D, a_1, A} \Pr(d_1, s_1, a_1) \log_2 \Pr(d_1 | s_1, a_1) - \\
 &\sum_{s_2, s_2, d_2, D, a_2, A} \Pr(d_2, s_2, a_2) \log_2 \Pr(d_2 | s_2, a_2) - \dots - \\
 &\sum_{s_L, s_L, d_L, D, a_L, A} \Pr(d_L, s_L, a_L) \log_2 \Pr(d_L | s_L, a_L) = \\
 &-\sum_{d, D, i=1}^k P_i(1) P_{id}(1) \log_2 P_{id}(1) - \\
 &\sum_{d, D, i=1}^k P_i(2) P_{id}(2) \log_2 P_{id}(2) - \dots - \\
 &\sum_{d, D, i=1}^k P_i(L) P_{id}(L) \log_2 P_{id}(L). \quad (20)
 \end{aligned}$$

根据式 (2), (6), (7) 和 (20), 可计算出确定性策略下不完全观测的马尔可夫决策过程的熵增加和可预测性增加, 而指标性熵测度 (如熵率和总可预测性增加等) 则没有定义

5 数值计算

下面论证熵测度的计算, 分析确定性策略下完全观测的马尔可夫决策过程 考虑到问题的一般性, 本文以出租车运行问题^[9]为例讨论马尔可夫决策过程 考虑出租车司机在 3 个城镇 (3 个状态) 游戏拉客时的决策问题, 可供选择的行动共有 3 种 出租车运行问题的数据, 即与状态和行为相关联的转移概率和报酬如表 1 所示

考虑确定性策略, 并假设系统运作开始时出租

表 1 出租车问题数据^[9]

状态 i	行动 a	转移概率 P_{ij}^a			报酬 w_{ij}^a			期望报酬 W_i^a	
		j	j	j	j	j	j		
1	1	1/2	1/4	1/4	10	4	8	8	
	2	1/16	3/4	3/16	8	2	4		2.75
	3	1/4	1/8	5/8	4	6	4		4.25
2	1	1/2	0	1/2	14	0	18	16	
	2	1/16	7/8	1/16	8	16	8		15
	3	0	0	0	0	0	0		0
3	1	1/4	1/4	1/2	10	2	8	7	
	2	1/8	3/4	1/8	6	4	2		4
	3	3/4	1/16	3/16	4	0	8		4.5

车司机处于城镇 C, 即初始状态的概率为 $P_0 = [0, 0, 1]^T$.

情形 1 采用值迭代方式, 根据表 1 可得到在不同决策时刻的策略为 $d_1 = [1, 1, 1]^T, d_2 = [1, 2, 2]^T, d_3 = [2, 2, 2]^T, d_4 = [2, 2, 2]^T, \dots$, 策略很快收敛到 $[2, 2, 2]^T$.

情形 2 假设策略是周期性的, 即 $d_1 = [1, 1, 1]^T, d_2 = [1, 2, 2]^T, d_3 = [2, 2, 2]^T, d_4 = [1, 1, 1]^T, d_5 = [1, 2, 2]^T, d_6 = [2, 2, 2]^T, \dots$ 根据第 3.2 节, 可计算出它们的熵测度, 其计算结果如图 2 和图 3 所示

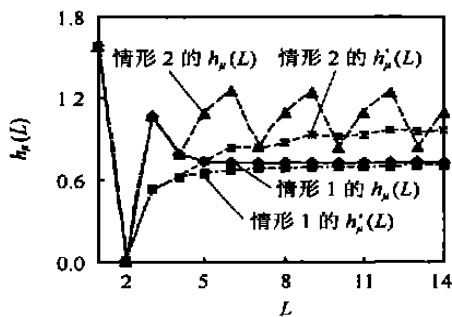


图 2 情形 1 和情形 2 的 $h_\mu(L)$ 和 $h_\nu(L)$

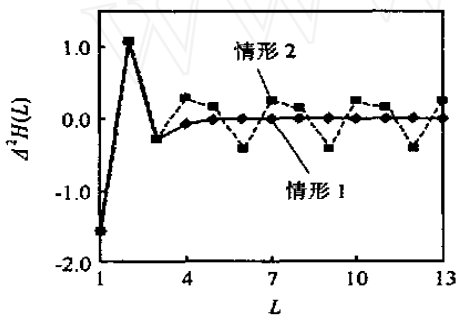


图 3 情形 1 和情形 2 的 $\Delta^2 H(L)$

对于情形 1, 因为策略收敛到一个特定的策略, 可得到其熵率为 0.72, 总可预测性增加为 -0.86 而 $h_\nu(L)$ 及 $h_\mu(L)$ 均收敛于 0.72, 可预测性增加 $\Delta^2 H(L)$ 很快收敛于 0 而在情形 2, 因为策略是周期性的, 所以 $h_\nu(L)$ 周期性地在 0.84, 1.09 和 1.25 之间摆动, $\Delta^2 H(L)$ 也周期性地在 -0.41, 0.25 和 0.16 之间摆动, $h_\mu(L)$ 则缓慢地收敛于 0.84, 1.09 和 1.25 的平均值 1.06 这样, 可将 1.06 看作熵率, 但其总可预测性增加则不能得到

对于情形 1, 因为策略收敛到一个特定的策略, 等同于策略不变的情况, 所以可计算所有的复杂性熵测度. 而情形 2 只能计算序列性的熵测度, 如

$h_\nu(L), h_\mu(L)$ 和 $\Delta^2 H(L)$ 的瞬时值, 而它们的收敛性是不可知的, 且指标性熵测度无法计算(熵率可以得到是特殊情况). 这是因为采用了周期性策略, 转移概率矩阵是变化的, 而且是周期性的

此数值仿真例子很好地证明了第 3.2 节的结论, 即指标性熵测度没有定义. 随着所采取策略的不同, 系统的复杂性也随之不同. 情形 1 实际上是收敛到一个特定策略, 而情形 2 是周期性策略, 因此情形 2 的决策问题比情形 1 复杂, 情形 2 的计算结果 $h_\nu(L)$ (0.84, 1.09 和 1.25) 始终大于情形 1 的 0.72, 这是符合实际的

6 结 论

本文用信息理论度量马尔可夫决策过程的复杂性, 在随机过程熵测度的基础上, 提出一套用于马尔可夫决策过程的基于信息理论的复杂性测度, 包括块熵、熵率、熵增加、可预测性增加等. 主要分析了在随机策略和确定性策略下完全观测和不完全观测马尔可夫决策过程的复杂性熵测度. 在随机策略下, 马尔可夫决策过程的熵测度与马尔可夫链的熵测度类似, 但在确定性策略下, 尤其在不同的决策时刻采用不同策略情况下, 只能计算序列性的熵测度, 而不能得到指标性的熵测度. 仿真结果也表明序列性的熵测度是波动的, 不能收敛于某一点

本文主要研究了马尔可夫决策过程的基于信息理论的复杂性测度方法. 这是一项全新的工作, 目前尚处于起步阶段, 许多方面需进一步研究: 1) 由于实际动态决策问题更加复杂, 其应用将更具挑战性; 2) 在不同策略下, 马尔可夫决策过程的熵测度是不同的, 揭示熵测度与策略之间的关系是一项值得关注的研究工作

参考文献(References):

- [1] Li W. On the relationship between complexity and entropy for Markov chains and regular languages[J]. *Complex Systems*, 1991, 5(4): 381-399.
- [2] Berthe V. Conditional entropy of some automatic sequences[J]. *J Physics A: Math Gen*, 1994, 27(24): 7993-8006.
- [3] Bialek W, Nemenman I, Tishby N. Predictability, complexity and learning [J]. *Neural Computation*, 2001, 13(11): 2409-2463.
- [4] Thomas M Cover, Joy A Thomas. *Elements of Information Theory* [M]. August: John Wiley & Sons, Inc, 1991.

(下转第 993 页)

- time-varying uncertainties [J]. *IEEE Trans on Automatic Control*, 1998, 43(2): 223-226
- [3] 苏宏业, 王景成, 褚健. 一类不确定动态系统的无记忆鲁棒镇定控制[J]. *自动化学报*, 1998, 24(4): 497-501.
(Su H Y, Wang J C, Chu J. Memoryless robust stabilization of a class of uncertain linear time-delay systems[J]. *Acta Automatica Sinica*, 1998, 24(4): 497-501.)
- [4] 费树岷, 姜偕富, 冯纯伯. 线性时滞系统对时滞参数的自适应控制[J]. *控制理论与应用*, 2001, 18(5): 686-690.
(Fei S M, Jiang X F, Feng C B. Adaptive control for linear delay systems to aim directly at delay parameter [J]. *Control Theory and Application*, 2001, 18(5): 686-690.)
- [5] Ge Jian-Hua, Frank P M, Lin Ching-Fang. H control via output feedback for state delayed systems[J]. *Int J Control*, 1996, 64(1): 1-7.
- [6] Wang J C, Su H Y, Chu J, et al. Robust H output feedback controller design for linear time-varying uncertain systems with delayed state [J]. *Control Theory and Application*, 1999, 16(3): 334-338
- [7] Su H Y, Wang J C, Chu J. Output feedback stabilizing controller for linear time-varying uncertain systems with delayed state[J]. *Control Theory and Application*, 1998, 15(6): 939-944
- [8] Zidong W, Biao H, Unbehauen H. Robust H observer design of linear time-delay systems with parametric uncertainty[J]. *Systems & Control Letters*, 2001, 42: 303-311.
- [9] 关新平, 林志云, 段广仁. 时变时滞不确定系统基于观测器的鲁棒控制器设计[J]. *信息与控制*, 1999, 28(3): 161-167.
(Guan X P, Lin Z Y, Duan G R. Design of observer-based robust controllers for uncertain systems with time-varying delay[J]. *Information and Control*, 1999, 28(3): 161-167.)
- [10] Han H C, Myung J C. Observer-based H controller design for state delayed linear systems [J]. *Automatica*, 1996, 32: 1073-1075
- [11] Petersen IR, Hollot Christopher V. A Riccati equation to the stabilization of uncertain linear systems [J]. *Automatica*, 1986, 22(4): 397-411.

(上接第 982 页)

- [5] Iwasaki T, Skelton R E, Gerome I J C. Linear quadratic suboptimal control with static output feedback[J]. *Syst Contr Lett*, 1994, 23(6): 421-430
- [6] Cao Y Y, Lam J, Sun Y X. Static output feedback stabilization: An LMI approach [J]. *Automatica*, 1998, 34(12): 1641-1645
- [7] Prempain E, Postlethwaite I. Static output feedback stabilization with H performance for a class of plants [J]. *Syst Contr Lett*, 2001, 43(3): 159-166
- [8] Gahinet P, Nemirovski A, Laub A J, et al. *LM I Control Toolbox* [M]. The Math Works Inc, 1995
- [9] Boyd S, El Ghaoui L, Feron E, et al. Linear matrix inequalities in system and control theory [A]. *SIAM Studies in Applied Mathematics* [C]. Philadelphia, SIAM, 1994
- [10] Davison E J, Wang S H. On pole assignment in linear multivariable systems using output feedback [J]. *IEEE Trans on Automatic Control*, 1975, 20(4): 516-518
- [11] Kimura H. Pole assignment by gain output feedback [J]. *IEEE Trans on Automatic Control*, 1977, 20(4): 509-516

(上接第 987 页)

- [5] Crutchfield J P, Feldman D P. Regularities unseen, randomness observed: Levels of entropy convergence [EB/OL]. <http://www.santafe.edu/sfi/publications/working/01-02-012.pdf>, 2001-02-12
- [6] Schümann T, Grassberger P. Entropy estimation of symbol sequences[J]. *Chaos*, 1996, 6(3): 414-427.
- [7] Frizelle G, Woodcock E. Measuring complexity as an aid to developing operational strategy [J]. *Int J of Operation and Production Management*, 1995, 15(5): 26-39
- [8] Puterman M L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming* [M]. John Wiley & Sons, Inc, 1994
- [9] Howard R A. *Dynamic Programming and Markov Processes* [M]. Cambridge: The Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons, Inc, 1960