第22卷 第1期
1996 年 1 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 22, No.1
Jan., 1996

短 文

用插值的粗粒化 Box Model 进行学习控制[1)]

刘 斌
(中国科学院系统所 100080)

张承福
(北京大学物理系 100871)

摘 要

针对一类未知动力学方程的单输入系统，采用插值的粗粒化 Box Model 进行学习控制. 与传统的 Box Model 及基于随机元胞自动机的控制模式比较，本模型设计思想自然，算法简单. 计算量小 (常常小几个数量级)，训练时间短，对三个较有代表性的实例的仿真结果也是令人满意的.

关键词: 学习控制，Box Model 随机元胞自动机.

1 引言

反馈镇定问题是控制界的研究问题之一. 按对系统信息量了解的多少，控制方法分为两类: 一类是已知系统的动力学方程，这类问题已有较成熟的现代控制理论工具. 另一类是未知系统的动力学方程，此类问题的处理方法之一是按照行为科学进行处理的学习控制，即对系统通过学习获得足够的信息进而进行控制，此系统亦称学习控制系统.

学习控制是人工神经元网络等人工智能控制研究的主要领域之一[1]. 学习方式有导师学习 (supervised learning)，激励学习 (reinforced learning) 等[2]. 其中，在激励学习方式下，接收到的外部信息通常是激励信号 (最简单的激励信号就是奖与罚)，按照一定评估标准对系统评估，符合要求的指标就奖，否则就罚. 据此激励信号相应地增强 (维持) 或削弱 (改变) 控制作用以便适应未知的环境.

本文讨论的是单输入的、未知动力学方程的不稳定动态系统，通过激励学习方式来控制此系统. D. Michie 等用 Box Model 首先控制了倒立车摆[3]. Box Model 的作法是将状态空间划分成一系列超矩形体 (称 Boxes)，对落入某个 box 的状态采取一定的控制方式，明显不足是需要构造较多的 box 以达到较好的控制效果. 拿倒立车摆来说，

1) 国家自然科学基金与高校博士学科点专项科研基金资助项目.
本文于 1993 年 7 月 5 日收到

状态空间划分成 $10\times10\times10\times10=10^4$ 个 box，训练量大，而此划分仍显粗糙，太粗糙用 Box Model 控制有时会失败. A. G. Barto 等[4] 虽然只分成 $6\times3\times3\times3=162$ 个 boxes，但对每个 box 的训练采用的是一个神经元网络，计算量仍不小. Y. C. Lee 等[5, 6] 用一种随机元胞自动机来控制倒立车摆，使训练量降到了只训练 384 条规则，不足是随机扰动较大以及要求控制的初态速率几乎为零，这降低了控制要求. 本文提出的模型对倒立车摆控制只划分 $3\times3\times3\times3=81$ 个 boxes，且只对 256 个边界点进行训练，对非边界点采用线性插值构造控制，大大减少了计算量，计算表明只须训练好 120 条关键性规则就能达到较好的控制效果，且初始状态速率不为零. 网络设计简单，训练时间短. 我们还用此模型对模拟人学自行车模型和双倒摆模型进行了控制(将另文讨论)，均得到较满意的结果，因而模型有一定的普适性.

2 基本模型与训练方法

考虑单输入非线性动态系统，状态为 X，$X\in\{(x_1,\cdots,x_n)\,||x_i|\leqslant b_i,\ i=1,\cdots,n\}\subset R^n$，控制 $u\in R$，不妨设系统平衡态为 $X=\{0,\cdots,0)$，且此平衡态在 $u=0$ 时是 Ляпунов 意义下不稳定的. 控制前提是未知动力学方程 $\dot{X}=f(t,X,u)$；控制目的是构造状态反馈 $u=g(t,X)$，使系统平衡态 $(0,\cdots,0)$ 稳定.

下面采用激励学习方式来进行控制. 在此方式下，评估函数选作 $V(t,X)$，$V(t,X)$ 是正定的. 由于 $u=0$ 时系统在原点不稳定，因而选择 u 的标准是使 V 递减或者不如 $u=0$ 时 V 增加得快. 具体讲，设状态 X_0 对应 $V_0=V(t_0,X_0)$，在 $u=0$ 时，经 Δt，系统到达状态 X_0'，相应 $V_0'=V(t_0+\Delta t,X_0')$，$\Delta V_0=V_0'-V_0$；在 $u=g(t,X_0)$ 时，经 Δt 系统到达 X_1，$V_1=V(t_0+\Delta t,X_1)$，$\Delta V_0'=V_1-V_0$，据选择 u 的标准，考虑 V 的净增量 $\Delta V=\Delta V_0'-\Delta V_0$，若 $\Delta V\leqslant0$，输出激励信号为奖，继续增加 u；若 $\Delta V>0$，输出激励信号为罚，减小 u；u 的改变值由 (1) 式决定.

$$\Delta u=-\eta\cdot\Delta V,\ \eta\text{ 为学习率.}\tag{1}$$

V 通常取作关于 X 的各个分量的正定二次型. 我们构造 $u=g(t,X)$ 的过程如下：首先将状态 X 的分量 x_i 所在区间 $[-b_i,b_i]$ 分成 (N_i-1) 等份，整个状态区域分成 $\prod_{i=1}^{n}(N_i-1)$ 个 boxes，为讨论方便引入 n 个线性变换将 $[-b_i,b_i]$ 变到 $[1,N_i]$. 这样任一状态 X 对应状态 $A=\{A_1,\cdots,A_n)$，$A_i\in[1,N_i]$，$i=1,\cdots,n$，状态 A 所在区域也被合成一系列 boxes，它们构成一网格，共有 $\prod_{i=1}^{n}N_i$ 个网格点 (即 A_i 均取整数的点). 对非网格点 A 的分量 A_i，$1<A_i<N_i$，与其相邻的整数值为 $I_i^-=[A_i]$，$I_i^+=[A_i]+1$，I_i^-，I_i^+ 相对 A_i 重要性的权重取为 $p_i^-=[A_i]+1-A_i$，$p_i^+=A_i-[A_i]$. 与 A 相邻的至多有 2^n 个网格点 $(I_1,\cdots,I_n)$，$(I_i=I_i^-$ 或 $I_i^+)$，采用如下方式对网格点编号：

$$M_i=I_1+N_1\cdot(I_2+N_2\cdot(I_3+\cdots(I_{n-1}+N_{n-1}\cdot I_n)\cdots)),\qquad i=1,\cdots,2^n,\tag{2}$$

与 A 相邻的第 M_i 个网格点使用的权重为

$$W_{M_i}=P_1\cdot P_2\cdots P_n,\qquad P_j=p_j^-\text{ 或 }p_j^+,\ i=1,\cdots,2^n. \tag{3}$$

我们称在网格点上所加的 u_i, $i=1,\cdots,N_1\times\cdots\times N_n$ 为一系列规则值, 一旦知道 u_i, 对非网格点 A 所加的 u 采用线性插值公式

$$u=\sum_{i=1}^{2^n}u_{M_i}W_{M_i}. \tag{4}$$

因而, 我们首先训练出规则值 u_i, $i=1,\cdots,N_1\times\cdots\times N_n$, 然后利用 (2)—(4) 构造非网格点上所加控制 u, 从而整个状态区域所加的 u 构造完毕. 训练分两种方案, 一是利用 (1) 单独训练每个 u_i, 二是利用 (1)—(4) 同时训练数个 u_i. 仿真结果如下.

3　对倒立车摆的仿真与研究

采用 [6] 中的倒立车摆模型, 描述如下:

l: 杆长, m: 杆质量, M: 车质量, g: 重力加速度, θ: 摆角, $\dot\theta$ 为角速度, Z 为位移, $\dot Z$ 为车速度. 系统状态变量为 $(\theta,\dot\theta,Z,\dot Z)$, 控制为所加的力 f. 仿真采用的动力学方程为

$$\ddot\theta=\frac{g\sin\theta-\cos\theta[(f+ml^2\dot\theta^2\sin\theta)/(M+m)]}{l\cdot[4/3-m\cos^2\theta/(M+m)]}$$

$$\ddot Z=\frac{f+ml(\dot\theta\sin\theta-\ddot\theta\cos\theta)}{M+m}$$

具体参数为 $M=1$ kg, $m=0.1$ kg, $g=9.8$ m/s^2, $l=1$ m, $|\theta|_{\max}=0.3$ rad, $|Z|_{\max}=1$ m, $|\dot\theta|_{\max}=0.3$ 1/s, $|\dot Z|_{\max}=1$ m/s, $|f|_{\max}=30(M+m)g$. 采用上面模型, 划分 $3\times3\times3\times3=81$ 个 boxes, 共 256 条规则, 评估函数为 $E=0.5(\theta^2+Z^2+\theta Z)+(\dot\theta^2+\dot Z^2+\dot\theta\dot Z)$, 规则值为 f_i, $i=1,\cdots,256$, 公式 (1)—(4)变为

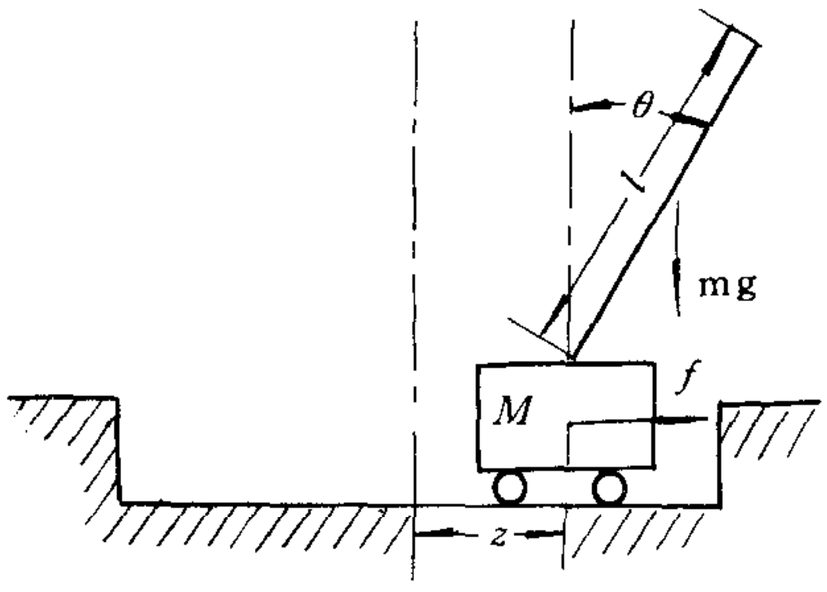

图 1　倒立车摆

$$\Delta f=-\eta\cdot\Delta E, \tag{5}$$

$$M_i=I_1+4(I_2+4(I_3+4I_4)),\quad i=1,\cdots,16, \tag{6}$$

$$W_{M_i}=P_1\cdot P_2\cdot P_3\cdot P_4,\ P_j=p_j^-\text{ 或 }p_j^+,\ i=1,\cdots,16,\ \ j=1,\cdots,4, \tag{7}$$

$$f=\sum_{i=1}^{16}f_{M_i}\cdot W_{M_i}. \tag{8}$$

训练方案有两种, 一是先单独训练出 f_i 再插值用于实时控制; 二是综合训练出数条 f_i 再用于实时控制.

3.1 方案一的仿真结果

采用 (5) 分别训练 f_i, $i=1, \cdots, 256$, 再用 (6)—(8) 构造整个区域上的 f 来控制. 好处是训练时间短. 效果如图 2、图 4. 图 4 为成功比率 u_α 与初始偏差范围 α 关系图 (即对 $\alpha(0\leqslant\alpha\leqslant 1.0)$, 在 $[-\alpha|\theta|_{\max}, \alpha|\theta|_{\max}]\times[-\alpha|\dot{\theta}|_{\max}, \alpha|\dot{\theta}|_{\max}]\times[-\alpha|Z|_{\max}, \alpha|Z|_{\max}]\times[-\alpha|\dot{Z}|_{\max}, \alpha|\dot{Z}|_{\max}]$ 内随机置点, 观其成功比率 u_α) 从图 4 可看出, 在 α 较小时, 成功比率 u_α 较高, α 越大, 控制越难.

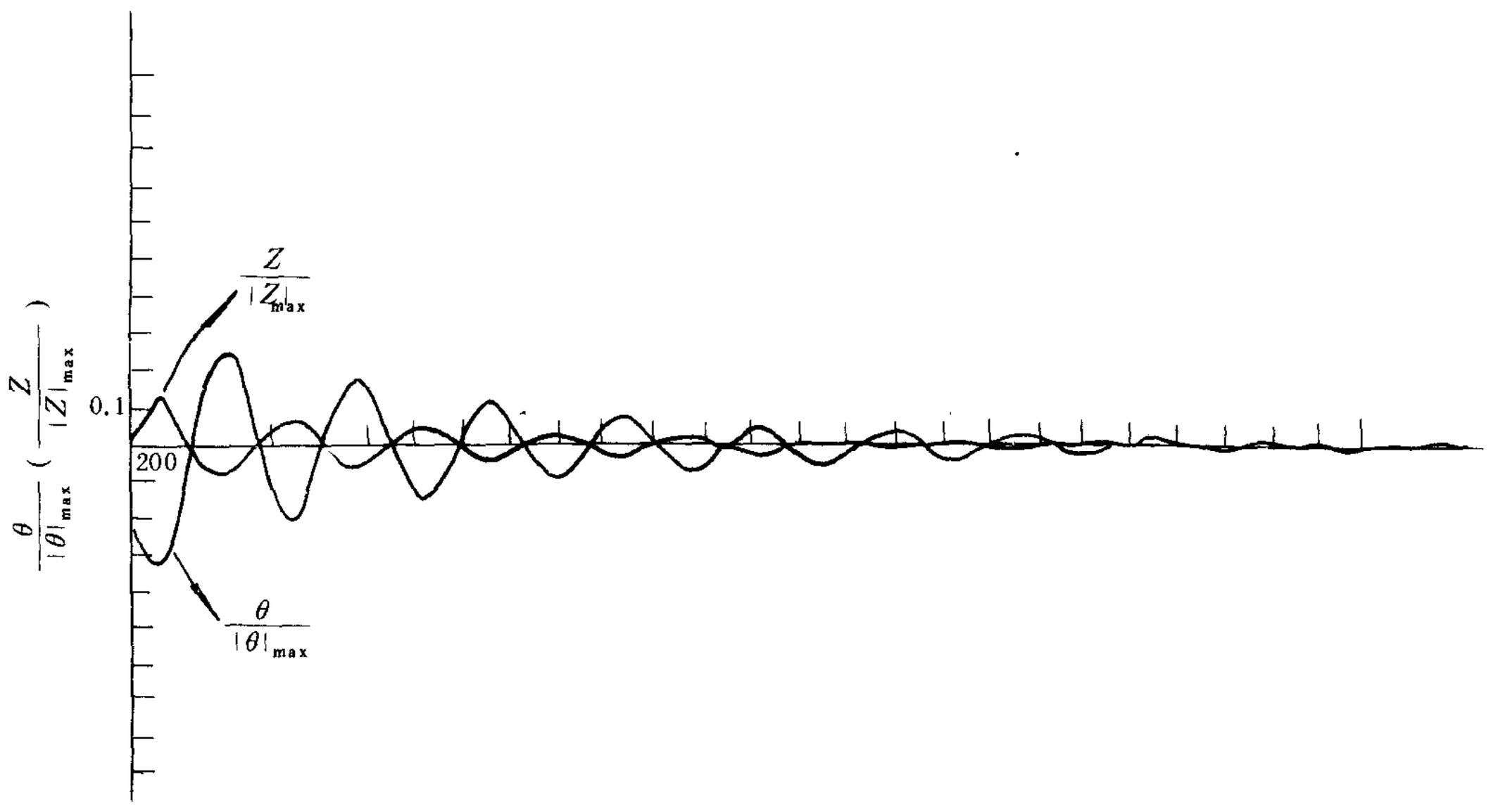

图 2 某一初态下的典型轨迹图 (方案一)

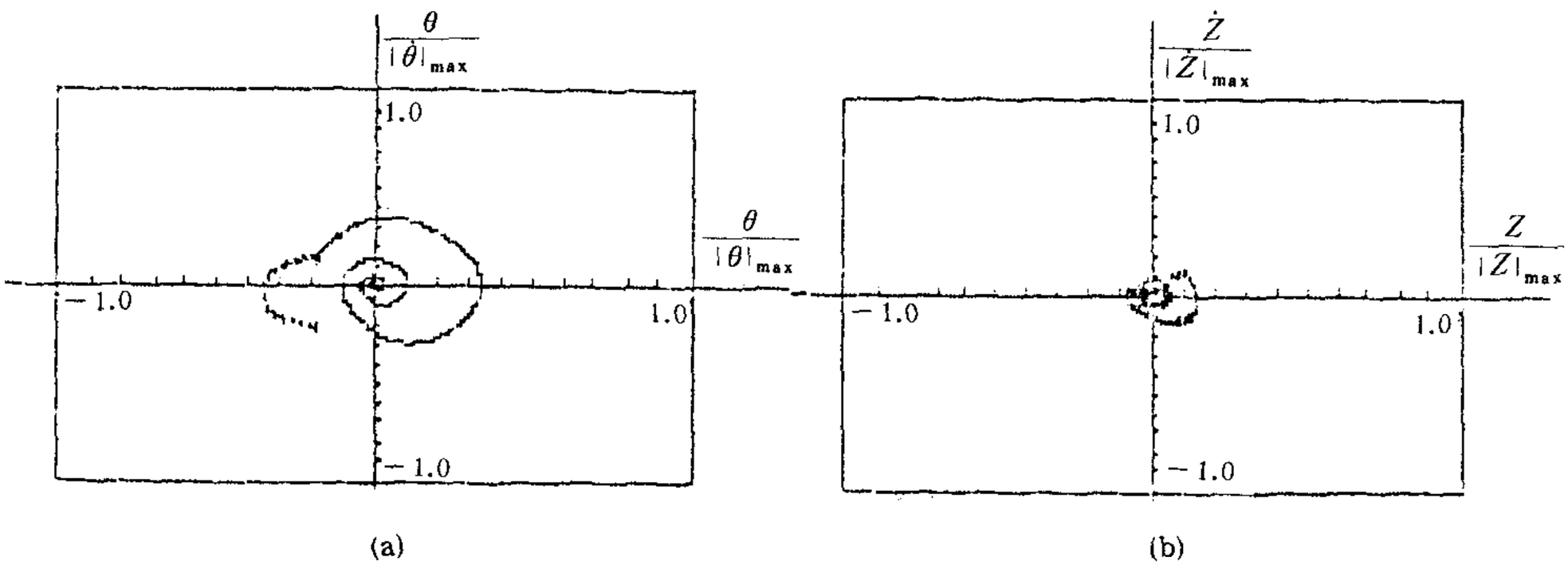

图 3 某一初态下的相图 (方案一)

3.2 方案二的仿真结果

前面的评估过程是从同一 X_0 出发不停地拉回到 X_0 来训练规则, 实时控制中不可能多次拉回到 X_0, 故评估修改如下: 分段加力分段对 X_1 评估, 前面采用 $\Delta E=(E_2-E_1)-$

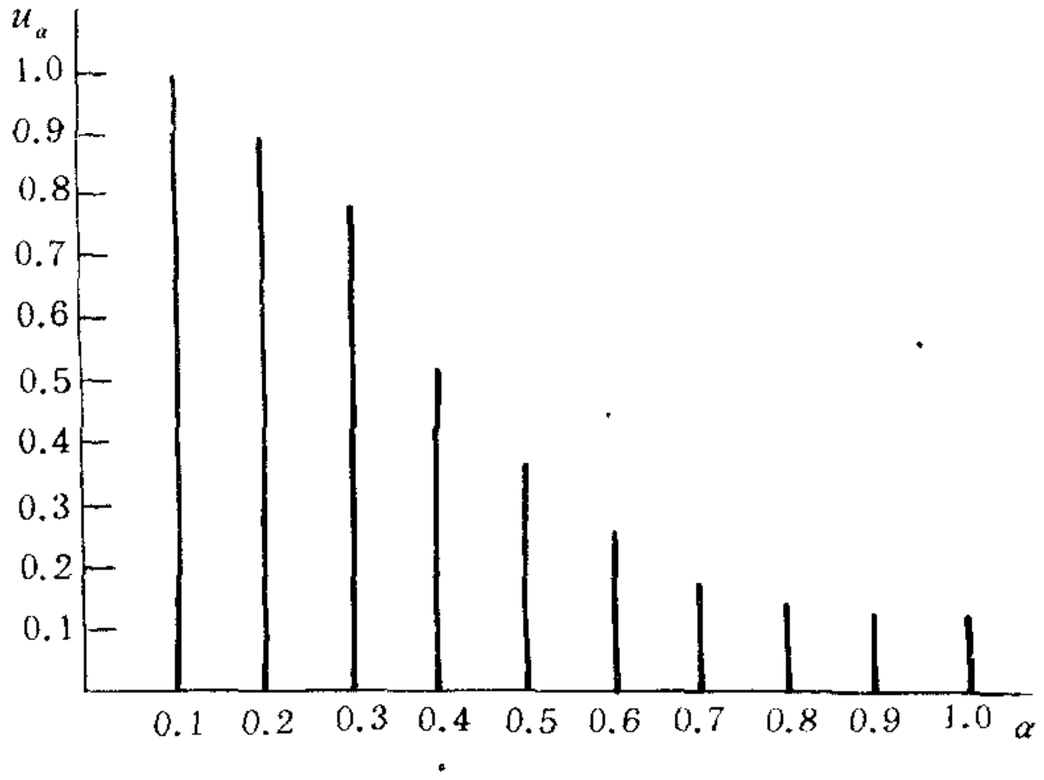

图 4　成功比率 u_α 与初始偏差范围 α 关系图 (方案一)

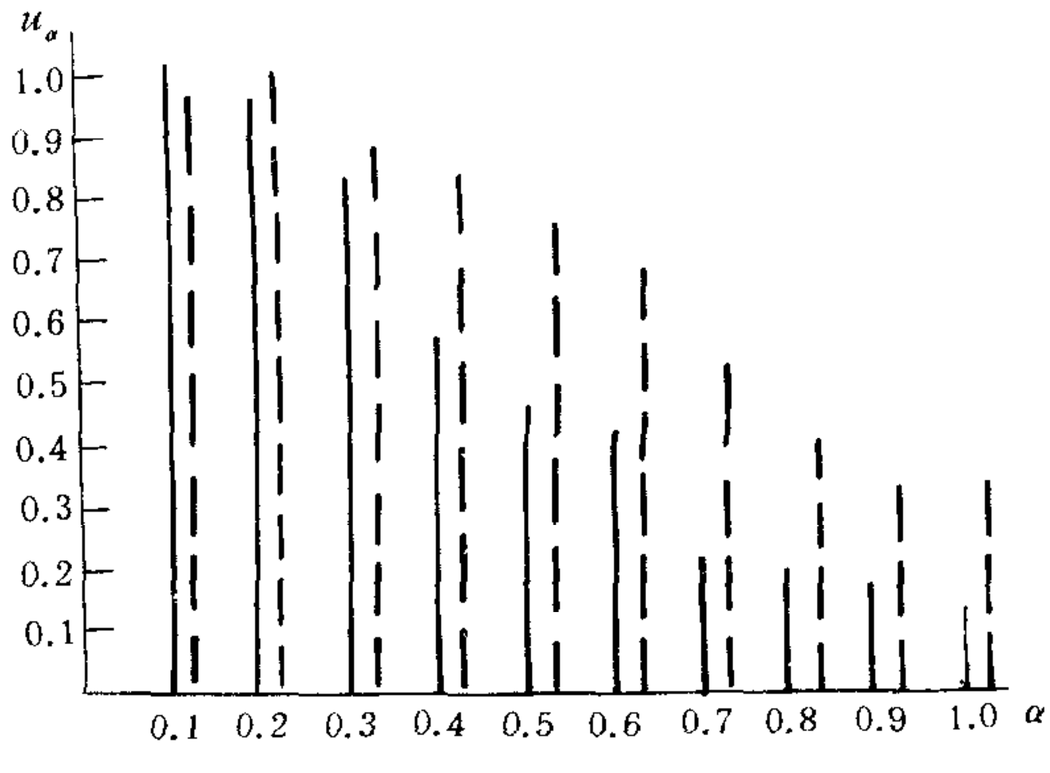

图 5　u_α 与 α 关系图 (方案二)
(虚线为初始速率为零的结果)

$(E_1'-E_1)$，现在为 $\Delta E^*=(E_2-E_1)-(E_1-E_0)$，这样任置初态，同时训练数条规则，(5) 改为 $\Delta f_{M_i}=-W_{M_i}\cdot\eta\cdot\Delta E^*$，$i=1,2,\cdots,16$，训练与插值不分开．为克服不可能同时训练好所有规则的困难，采用同时训练逐步自动关闭训练的措施．具体讲，对第 i 条规则使用的频数加权累加，若达到规定的某值就关闭对它的训练．控制效果如图 5.

从图 5 来看，在 $\alpha\leqslant 0.3$ 时，$u_\alpha\geqslant 70\%$，说明结果是较好的．按文[6] 中仿真要求，使初始 $\dot{\theta}$, $\dot{Z}$ 为零得到图 5 虚线所示结果，可看出，u_α 明显提高，在 $\alpha=1.0$ 时，也有 30% 成功率，说明文 [6] 仿真要求较低，同时可看出初始速率偏差对控制结果影响较大．另外频数统计表明经常使用的仅有 120 条关键规则，训练好这些规则就能达到一定的控制效果.

3.3　讨论

(1) 评估函数的选取是关系到成败的重要一环.

(2) 计算表明规则初始值对最终训练结果无影响，但在有限时间里，还是要注意 η 的选取.

(3) 同时训练逐步自动关闭的方式可避免训练过分问题，但可能使某些规则训练不充分.

4　结论与展望

采用本模型我们还对模拟人学自行车模型及双倒摆模型进行了控制 (另文讨论)，均得到满意的结果，说明本模型具有一定的普适性．本模型划分的 boxes 是最粗糙的，用 Box Model 控制会失败，而我们可控制到相当好的结果．网络构造简单，训练时短，计算量大为减小 (通常小几个量级)，与文 [6] 相比，模型的扰动较小．值得进一步研究的是：① 如何自动地调试出评估函数，② 推广到多输入系统，③ 非线性插值是否有更好的效果.

参 考 文 献

[1] 张承福. 神经网络系统. 力学进展, 1988, **18**(2): 145—160.

[2] Lee Y C. Evolution. Learning and cognition, World Scientific, Singapore, 1988.

[3] Michie D, et al. Boxes, an experiment in adaptive control, in machine Intelligence 2, E. Dale and D. Michie, eds. (Oliver and Boyd, Edinburgh, 1968), 137—152.

[4] A G Barto, et al. Neuralike adaptive elements that can solve difficult learning control problem. IEEE Trans. on S. M. C., 1983, **SMC-134**: 834—846.

[5] Lee Y C, et al. Adaptive stochastic cellular automata: theory. Physica D, 1990, **45**: 159—180.

[6] Lee Y C, et al. Adaptive stochastic cellular automata: Application. Physica D, 1990, **45**: 181—188.

LEARNING CONTROL BY COARSE-GRAINED BOX MODEL WITH INTERPOLATION

LIU BIN

(*Institute of Systems Science, Academia Sinica* 100080)

ZHANG CHENGFU

(*Department of Physics, Peking University* 100871)

ABSTRACT

In this paper, learning control is applied to a single-input system with unknown dynamical equations, using coarse-grained Box Model with interpolation. As compared with the traditional Box Model or with Stochastic Cellular Automata, this model is more reasonal in idea, much simpler in algorithm, less in computation amount and less training time. Three illustrative simulation examples are given and show satisfactory results.

Key words: Learning control. box model, stochastic cellular automata.