DOI:10.16185/j.jxatu.edu.cn.1990.02.006

第10卷 第2期 西安工业学院学报 Vol.10 No.2
1990年4月 JOURNAL OF XIAN INSTITUTE OF TECHNOLOGY Apr. 1990

一类概率指标动态规划模型

孙明轩 陈阳泉
（计算机控制教研室）

摘　要

本文提出了一个以氧气转炉炼钢控制过程为工程背景的随机控制模型。给出了它的动态规划方程。对于一类简单情形，其最优策略可通过求解一非线性方程组得到。

关键词：动态规划，随机控制。

1　模型

工业生产中的某些最优控制过程，可以归结为具有概率指标的随机控制问题。例如，冶金工业中的转炉炼钢过程[3,4]，熔炼目的是，使冶炼终了时钢水的温度和含碳量等，同时达到钢种所要求的指标范围。氧气转炉需在几十分钟内完成造渣、脱硫、脱磷、脱碳、去除非金属夹杂及升温等基本任务。冶炼机理复杂。整个冶炼过程属于高温操作，连续测温困难。影响钢水终温和终碳的随机因素很多，诸如吹炼时间、枪位、氧气压力流量、副料、间隙时间以及炉衬变化等。因此，为了提高终点命中率，生产现场便采用副枪进行动态测试的生产工艺。在吹炼开始时，根据铁水成份、温度及要求的钢水终碳终温，控制吹氧时间和副料加入量；吹炼后期降下副枪，测量钢水含碳量和温度。若主吹未命中，则根据副枪获得的信息，控制补吹时间和副料加入量，保证钢水达到钢种要求。实际上，上述是这样一类控制问题：受控过程按段演变，在每一段对于受控过程所施加的控制，应使得过程尽可能地命中所希望的目标区域（我们称之为靶集）。在本文中，假设这类受控过程的状态只可能处于两类区域：一为正常的状态区域 B，状态进入此域可继续转移；另一为希望命中的区域 A，状态一旦进入此域，过程就停止转移。

我们给出上述问题的模型[1]如下

受控过程的状态转移律

$$\boldsymbol{x}_{k+1}=\boldsymbol{f}_k(\boldsymbol{x}_k,\ \boldsymbol{u}_k,\ \boldsymbol{w}_k)\qquad k=0,1,\cdots,N-1 \tag{1}$$

1988—12—14 收到

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

其中，状态 $x_k \in S_k$，控制 $u_k \in C_k$；当 u_k 受到约束时，对于所有 $x_k \in S_k, u_k \in U_k(x_k) \sqsubset C_k$，$U_k(x_k)$是第 k 段允许控制集。随机干扰 $w_k \in D_k$，其分布函数只依赖于x_k，u_k，而与以前的干扰w_{k-1}，w_{k-2}，…，w_0无关。初态 $x_0 \in S_0$的概率分布 $F(x_0)$给定。

该受控过程在第 i 段的停止条件是 $x_i \in A$。一般地，$S_0 \cap A = \phi$。这是因为当 $x_0 \in A$ 时，不需要对过程施加控制。容易看出，受控过程的状态具有无后效性。即

$$F(x_k | x_{k-1}, x_{k-2}, \cdots, x_0) = F(x_k | x_{k-1}) \tag{2}$$

指标函数

$$J_1 = \sum_{k=1}^{N} P_r(x_0 \in S_0, x_1 \in B, x_2 \in B, \cdots, x_{k-1} \in B, x_k \in A) \tag{3}$$

这时问题归结为在允许控制集中，寻找一允许控制策略$\{u_0^*(x_0), u_1^*(x_1), \cdots, u_{N-1}^*(x_{N-1})\}$最大化 J_1。为叙述方便，文中称由 J_1 构成的问题为 PSDP。

历程N为有限数。氧气转炉炼钢中 $N=2$。炉长以一次补吹命中的百分数作为命中率的衡量标准，这一数值被称为终点命中率。N具体取多少应视实际情况而定。通常，生产过程在一定段数的控制作用之后，不命中的概率已很小。例如对于PSDP，记

$$q_k \triangleq p_r(x_0 \in S_0, x_1 \in B, x_2 \in B, \cdots, x_k \in B)$$
$$Q_{k-1} \triangleq p_r(x_k \in B | x_{k-1} \in B)$$

则有

$$q_k = q_1 \prod_{j=1}^{k} Q_j \tag{4}$$

因此，$\lim\limits_{k \to \infty} q_k = 0$。这说明当 $k \to \infty$时，过程终点状态必然命中。

国外曾有人建立了转炉冶炼过程中钢水温度和含碳量的离散状态方程，并提出用动态规划方法求解最优控制。文献[2]以及文献[5]，[6]，分别深入地研究了以马尔可夫链和随机微分方程描述的概率指标的命中控制问题。本文试图借鉴前人的工作，为实际应用做些初步的理论准备。

2 求解公式

首先，应用动态规划方法给出 PSDP 的求解公式，即动态规划基本方程。下面先引入要用到的一些结果。

引理 1 假设对于 $x \in S, u \in U$，$f: S \times U \to R$，$f(x,u)$ 关于 u 存在最大值 $f(x,u^*)$，关于 x 在 $\Delta \sqsubset S$ 上可积，则有

$$\max_{u(x)} \int_{\Delta} f(x,u) \mathrm{d}x = \int_{\Delta} \max_{u} f(x,u) \mathrm{d}x \tag{5}$$

证 对于所有的容许控制 u，我们有

$$f(x, u) \leqslant f(x,u^*) = \max_{u} f(x,u)$$

从而

$$\int_{\Delta} f(x,u) \mathrm{d}x \leqslant \int_{\Delta} f(x,u^*) \mathrm{d}x$$

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

将上式左边关于所有的容许控制取最大值，得

$$\max_{u(x)}\int_{\triangle}f(x,u)\,\mathrm{d}x\leqslant\int_{\triangle}f(x,u^*)\,\mathrm{d}x \tag{6}$$

因为 u^* 也是容许控制，所以有

$$\int_{\triangle}f(x,u^*)\,\mathrm{d}x\leqslant\max_{u(x)}\int_{\triangle}f(x,u)\,\mathrm{d}x \tag{7}$$

结合（6）和（7）式，得证。

引理 2　假设 X,Y 为随机变量，对于随机事件 A,B，有

$$p_r(X\in A,\ Y\in B)=\int_A p_r(Y\in B|X=x)\,\mathrm{dF}(x)$$

其中 $\mathrm{F}(x)$ 为 X 的分布函数。

证明由条件概率的定义易得。

引理 3　PSDP 的指标函数满足递推公式

$$V_k(x_k)=p_r(x_{k+1}\in A|x_k)+\int_B V_{k+1}(x_{k+1})\,\mathrm{dF}(x_{k+1}|x_k)$$

$$k=0,1,\cdots,N-1$$

$$V_N(x_N)=0$$

证　由引理 2 和状态的无后效性，有

$$J_1=\sum_{k=1}^{N}p_r(x_0\in S_0,x_1\in B,x_2\in B,\cdots,x_{k-1}\in B,x_k\in A)$$

$$=\int_{S_0}\Big\{p_r(x_1\in A|x_0)+\int_B\Big[p_r(x_2\in A|x_1)+\cdots+\int_B\Big(p_r(x_{N-1}\in A|x_{N-2})$$

$$+\int_B p_r(x_N\in A|x_{N-1})\,\mathrm{dF}(x_{N-1}|x_{N-2})\Big)\mathrm{dF}(x_{N-2}|x_{N-3})\cdots\Big]\mathrm{dF}(x_1|x_0)\Big\}\,\mathrm{dF}(x)$$

我们定义

$$V_k(x_k)=p_r(x_{k+1}\in A|x_k)$$

$$+\int_B\Big[p_r(x_{k+2}\in A|x_{k+1})+\cdots+\int_B\Big(p_r(x_{N-1}\in A|x_{N-2})$$

$$+\int_B p_r(x_N\in A|x_{N-1})\,\mathrm{dF}(x_{N-1}|x_{N-2})\Big)\mathrm{dF}(x_{k+2}|x_{N-3})\cdots\Big]\mathrm{dF}(x_{k+1}|x_k)$$

由此定义易得结论。

命题 1　对于 PSDP，$\{u_0^*(x_0),u_1^*(x_1),\cdots,u_{N-1}^*(x_{N-1})\}$ 为最优策略的充分必要条件是它满足方程

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$J_k(\boldsymbol{x}_k)=\max_{\boldsymbol{u}_k(\boldsymbol{x}_k)}\left\{p_r(\boldsymbol{x}_{k+1}\in A|\boldsymbol{x}_k)+\int_B J_{k+1}(\boldsymbol{x}_{k+1})\mathrm{dF}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\right\} \tag{8}$$

$$k=0,1,\cdots,N-1$$

$$J_N(\boldsymbol{x}_N)=0$$

证 只须对子过程 $[k,N]$ 作出证明，其余可以类推。先证必要性，设 $p(k,N)$ 为 $[k,N]$ 上的策略，由引理 1 及引理 3，得

$$\max_{p_{(k,N)}}V_k(\boldsymbol{x}_k)=\max_{\boldsymbol{u}_k(\boldsymbol{x}_k)}\max_{p_{(k+1,N)}}\left\{p_r(\boldsymbol{x}_{k+1}\in A|\boldsymbol{x}_k)+\int_B V_{k+1}(\boldsymbol{x}_{k+1})\mathrm{dF}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\right\}$$

$$=\max_{\boldsymbol{u}_k(\boldsymbol{x}_k)}\left\{p_r(\boldsymbol{x}_{k+1}\in A|\boldsymbol{x}_k)+\int_B\max_{p_{(k+1,N)}}V_{k+1}(\boldsymbol{x}_{k+1})\mathrm{dF}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\right\}$$

我们定义 $J_k(\boldsymbol{x}_k)=\max\limits_{p_{(k,N)}}V_k(\boldsymbol{x}_k)$，则得（8）式。

充分性。任选容许策略 $p(k,N)$，有

$$V_k(\boldsymbol{x}_k)=p_r(\boldsymbol{x}_{k+1}\in A|\boldsymbol{x}_k)+\int_B V_{k+1}(\boldsymbol{x}_{k+1})\mathrm{dF}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$$

$$\leqq p_r(\boldsymbol{x}_{k+1}\in A|\boldsymbol{x}_k)+\int_B\max_{p_{(+1,N)}}V_{k+1}(\boldsymbol{x}_{k+1})\mathrm{dF}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$$

$$\leqq\max_{\boldsymbol{u}_k(\boldsymbol{x}_k)}\left(p_r(\boldsymbol{x}_{k+1}\in A|\boldsymbol{x}_k)+\int_B\max_{p_{(k+1,N)}}V_{k+1}(\boldsymbol{x}_{k+1})\mathrm{dF}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\right)$$

由于 $p_{(k,N)}$ 的任意性，而 $p^*_{(k,N)}$ 又满足（8）式，故得证。

在推证命题 1 的过程中，对于靶集 A 和 $\boldsymbol{w}_k$ 的分布形式未做特别假定，因此命题 1 的适用范围是较广泛的。这时求解方程（8），一般是无法得到解析解的。这些公式的数值计算可以按动态规划的常规算法在计算机上进行。但是，从实际问题出发，可将原问题简化进而得到较简单的（甚至得到易于进行实时控制的）求解公式，而简化后的问题又具有较大的实际适应范围。这些在第 3 节中讨论。

3 简单情形

我们考虑状态转移律为

$$\boldsymbol{x}_{k+1}=\boldsymbol{f}_k(\boldsymbol{x}_k,\boldsymbol{u}_k)+\boldsymbol{w}_k,\qquad k=0,1,\cdots,N-1 \tag{9}$$

$\boldsymbol{u}_k$ 无约束，$\boldsymbol{w}_k$ 的分布与 $\boldsymbol{x}_k$，$\boldsymbol{u}_k$ 无关，并且靶集为如下的偏差带，通常的工业生产过程的指标要求具有这种形式

$$\mathrm{A}\triangleq\{\boldsymbol{x}\,|\,\boldsymbol{a}\leqq\boldsymbol{x}\leqq\boldsymbol{b}\} \tag{10}$$

其中，$\boldsymbol{a}$ 为向量 $\boldsymbol{x}$ 的下限向量，$\boldsymbol{b}$ 为上限向量。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

命题 2 考虑满足上述条件的 PSDP，对于固定的 k，如果指标函数

$$p_r(\boldsymbol{a} \leqq \boldsymbol{x}_{k+1} \leqq \boldsymbol{b} \mid \boldsymbol{x}_k) \tag{11}$$

当 $\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{u}_k) = \boldsymbol{f}_k^*$ 时达到最大值，那么 PSDP 的最优策略可通过求解下述方程组得到

$$\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{u}_k^*(\boldsymbol{x}_k)) = \boldsymbol{f}_k^*, \quad k = 0, 1, \cdots, N-1 \tag{12}$$

证 对于方程（8），考查 $k = N-1$ 时

$$J_{N-1}(\boldsymbol{x}_{N-1}) = \max_{u_{N-1}} p_r(\boldsymbol{a} \leqq \boldsymbol{x}_N \leqq b \mid \boldsymbol{x}_{N-1})$$

由假设知 $\boldsymbol{f}_{N-1} = \boldsymbol{f}_{N-1}^*$ 时，$J_{N-1}(\boldsymbol{x}_{N-1})$ 达到最大值，且它与 $\boldsymbol{x}_{N-1}$ 无关。

$k = i$ 时，有

$$J_i(\boldsymbol{x}_i) = J_{i+1}(\boldsymbol{x}_{i+1}) + \left(1 - J_{i+1}(\boldsymbol{x}_{i+1})\right)\left(\max_{u_i} p_r(\boldsymbol{a} \leqq \boldsymbol{x}_{i+1} \leqq \boldsymbol{b} \mid \boldsymbol{x}_i)\right)$$

由假设及逆向递推可证得结论。

注 1 命题 2 未指出 $\boldsymbol{w}_k$ 服从分布的形式。

注 2 由该命题知，这时的 PSDP 的最优策略可以通过两步求得：①求解(11)式问题的 $\boldsymbol{f}_k^*$；②求解(12)式非线性方程组。这比求解方程(8)所占机时少得多，并且不占内存。这时的控制问题是一系列静态最优化问题，从动态角度考虑并未使受控过程得到改善。这是原问题的一种退化情形。该结论反映了通常人脑对单历程控制的思维方式，即使得受控过程瞄准靶心，然后施加控制的方法。

注 3 对于线性系统

$$\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{u}_k) = \mathrm{A}_k \boldsymbol{x}_k + \mathrm{B}_k \boldsymbol{u}_k$$

最优控制应满足

$$\mathrm{B}_k \boldsymbol{u}_k = \boldsymbol{f}_k^* - \mathrm{A}_k \boldsymbol{x}_k$$

其中，$\boldsymbol{f}_k^*$ 为可离线确定的常向量。

注 4 对于一维情形的几类概率分布，求得的 f^* 见表 1。对于复杂的概率分布通过计算机寻优可求出 f^* 的数值解。

表 1 一维 f^* 值

w 的密度函数	f^*
$p(x) = \begin{cases} \dfrac{1}{c-d} & c \leqslant x \leqslant d \\ 0 & \text{其它} \end{cases}$ $c < d$，常数	$a-c \leqslant f^* \leqslant b-d$
$p(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$ $\lambda > 0$，常数	$f^* = a$
$p(x) = \dfrac{1}{2\lambda} e^{-\frac{\lvert x-\mu \rvert}{\lambda}}$ $\lambda > 0, \mu$，常数	$f^* = \dfrac{a+b}{2} - \mu$

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

推论 在命题2的条件下，如果 $w_k \in R^n$ 服从正态分布 $N(\mu_k,\sigma_k)$，则有

$$f_k^* = \frac{a+b}{2} - \mu_k$$

证 考查下述指标函数的最大化问题

$$\max_{f_k} J = \int_{a-f_k}^{b-f_k} p(x)\,\mathrm{d}x \tag{13}$$

其中 $$p(x) = \frac{1}{(2\pi)^{n/2}|\sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_k)^T\sigma_k^{-1}(x-\mu_k)\right\}$$

μ_k 为 n 维列向量，σ_k 为 $n\times n$ 正定对称矩阵。我们应用矩阵论中的一个结果：若 σ_k 是正定对称矩阵，则存在非奇异矩阵 L，使 $\sigma_k = L^T L$。对(13)式作线性变换

$$y = L^{-1}(x-\mu_k)$$

该变换的雅可比行列式为 $|L| = |\sigma_k|^{1/2}$，有

$$\max_{f_k} J = \int_{L^{-1}(a-f_k-\mu_k)}^{L^{-1}(b-f_k-\mu_k)} \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}y^T y\right)\mathrm{d}y$$

对 J 关于 f_k 求偏导并令其为零，可证得结论。

参 考 文 献

［1］ 吴沧浦．最优控制的理论与方法．北京：国防工业出版社，1989．

［2］ 吴沧浦．可控马尔可夫链的一种最优决策．自动化学报，1964（3）

［3］［日］ 野坂康雄．钢铁工业中的计算机控制．上海市：上海科技出版社，1982。

［4］ 孙明轩．转炉补吹脱碳的统计规律．炼钢，1989（1）

［5］ Van Mallert, L. J. et al. Numerical Solution of an Optimal Control Problem with a Probability Criterion. IEEE Trans. Auto. Contral AC-17, 1972(3)

［6］ Reid, D. W. et al. Hitting a Target with Maximum Probability. Int. J. Systems. Sci. 1980, 11（9）

A DYNAMIC PROGRAMMING MODEL WITH A PROBABILITY CRITERION

Sun Mingxuan Chen Yangquan

Abstract

This paper presents a stochastic control model which is of practical interest, and gives the DP equation of the model. Under some simple conditions, the optimal control strategy is obtained by solving the nonlinear equations.

Key words: dynamic programming; stochastic control

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net