遗传算法种群多样性的分析研究*

张晓缋　戴冠中　徐乃平
(西北工业大学自动控制系·西安,710072)

摘要：种群的多样性是遗传算法进化的前提条件.本文提出用种群方差和熵两个量来全面刻画遗传算法中种群的多样性,分析了选择、交换和变异三个主要算子对种群方差和熵的影响,同时比较了编码机制对种群多样性的影响,得出了一些重要的结论.

关键词：遗传算法；选择操作；交换操作；变异操作；种群多样性

1　引　言

近年来虽然遗传算法得到了广泛的应用,并取得了一定的成果.但目前没有完整的理论解释算法的机理,阻碍了遗传算法的应用,甚至引起了部分人的怀疑.

Holland 提出的图式定理是遗传算法的主要理论,但图式定理只揭示了种群平均适应值的进化情况,对种群中个体的分布情况并未给出信息.我们知道种群只有在保持一定的多样性的情况下才能进化,那么种群的多样性用什么来度量?选择、交换和变异操作对种群的多样性又有什么样的影响?编码机制对种群多样性有什么影响?这些问题对进一步理解遗传算法是很重要的,这也是本文要解决的问题.

2　种群多样性的表示

我们可以从两方面来理解种群的多样性,一是种群的空间分布,显然从这个意义上来看,种群的空间分布越大,种群的多样性就越强.因此,我们用方差来描述种群的空间分布情况.

定义 1　若第 t 代种群中的个体 x_t^i 由 L 个基因构成,即 $x_t^i=[x_t^{i(1)}\,x_t^{i(2)}\cdots x_t^{i(L)}]$, $i\in\{1,2,\cdots,N\}$,定义第 t 代种群的平均个体 $\overline{x}_t$ 如下:

$$\overline{x}_t=[\overline{x}_t^{(1)}\quad \overline{x}_t^{(2)}\quad \cdots\quad \overline{x}_t^{(L)}]. \tag{1}$$

其中 $\overline{x}_t^{(l)}=\sum_{i=1}^{N}x_t^{i(l)}/N$,由此定义第 t 代种群的方差为

$$D_t=[D_t^{(1)}\quad D_t^{(2)}\quad \cdots\quad D_t^{(L)}]. \tag{2}$$

其中 $D_t^{(l)}=\sum_{i=1}^{N}(x_t^{i(l)}-\overline{x}_t^{(l)})^2/N$, $l\in\{1,2,\cdots,L\}$.

从定义 1 可以看出方差 D_t 是 L 维的行向量,每一个分量表示出了种群在这维坐标上的空间分布.显然 $D_t^{(l)}$ 越大,则种群在第 l 维坐标上的空间分布就越大.

方差仅反映出了种群的空间偏离程度,还不能完全刻画出种群的多样性.例如,种群{1,2,3,4,5}由五个个体构成,显然 $D=2$;而种群{1,1,1,1,5},方差 $D=2.56$.虽然后者比前者的方差大,但前者比后者更具有进化能力,为此我们引入描述种群多样性的第二个量——熵.

定义 2　若第 t 代种群有 Q 个子集:$S_{t1},S_{t2},\cdots,S_{tQ}$,各个子集所包含的个体数目记为 $|S_{t1}|,|S_{t2}|,\cdots,|S_{tQ}|$,且对任意 $p,q\in\{1,2,\cdots,Q\}$, $S_{tp}\cap S_{tq}=\varnothing$, $\bigcup_{q=1}^{Q}S_{tq}=A_t$, A_t 为第 t 代种群

* 国家自然科学基金重大项目资助课题(69391900).
本文于 1996 年 8 月 8 日收到. 1997 年 3 月 5 日收到修改稿.

的集合．则定义第 t 代种群的熵如下

$$E_t = -\sum_{j=1}^{Q} p_j \log(p_j). \tag{3}$$

其中 $p_j = \frac{|S_{tj}|}{N}$，N 为种群中个体数目．

定义2告诉我们两点，一是当种群中所有个体都相同时，即 $Q=1$，这时熵取最小值 $E=0$．第二点是当 $Q=N$ 时，熵取最大值 $E=\log(N)$．种群中个体类型越多，分配得越平均，熵就越大．对于十进制编码，熵的最大值为 $E^D_{\max}=\log(N)$；对二进制编码熵的最大值为 $E^B_{\max}=\log(\min(N,2^L))$．

如果种群的方差和熵都很大，种群的情况如图1(a)所示，一般初始种群就是这个样子．当种群的方差很大，熵很小时，如图1(b)所示，种群收敛于几个分散的点上．种群的方差很小，熵很大时，种群集中在一个很小的区域内，如图1(c)所示．种群的方差和熵都很小时，种群收敛，如图1(d)所示，经过很多代进化后，种群处于这种状态．

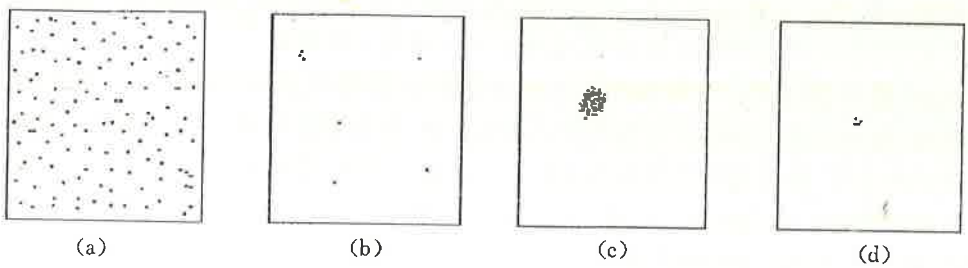


图1 种群多样性与方差和熵的关系

3 遗传算子对种群方差的影响

3.1 选择操作对种群方差的影响

遗传算法的选择操作可以描述为从个体 $x_t^1, x_t^2, \cdots, x_t^N$ 中选择出个体 $x_t'^1, x_t'^2, \cdots, x_t'^N$，满足概率 $P_r(x_t'^i = x_t^j | x_t^1, \cdots, x_t^N) = f(x_t^j) / \sum_{k=1}^{N} f(x_t^k)$，这里 $f(\cdot)$ 是适应函数[1,2,3]．

Xiaofeng Qi 在[4]中得出种群数目无穷情况下，选择操作对种群的概率密度函数 $g(\cdot)$ 的影响为

$$g_t'(x) = \frac{g_t(x)}{\mathrm{E}[f_t]}.$$

其中，E[·]表示数学期望，Xiaofeng Qi 还证明了当 $t \to \infty$ 时，种群概率密度函数 g_t 收敛于冲激函数

$$g_{t\to\infty} = \delta(x - x^*). \tag{4}$$

这里，x^* 是初始种群中适应值最好的个体．

定理1 种群数目为无穷，在只有选择操作作用下，有

$$\lim_{t\to\infty} D_t = 0. \tag{5}$$

证 由(4)式显然定理1成立．

定理1虽然说明了种群数目为无穷时，方差收敛于0，但并不保证方差单调下降．图2表示在只有选择操作作用下种群方差的变化情况．

3.2 交换操作对种群方差的影响

交换操作可以描述为随机地从种群 $x_t'^1, x_t'^2, \cdots, x_t'^N$ 中选出两个个体，依交换概率 P_c 进行基

因重组运算，产生种群 $x_t^{\prime\prime 1}, x_t^{\prime\prime 2}, \cdots, x_t^{\prime\prime N}$. Xiaofeng Qi 在[5]中证明了当种群数目无穷，交换操作保证每个坐标的边缘概率密度函数不变，即

$$g_{x_t^{\prime\prime(l)}}(x^{(l)}) = g_{x_t^{\prime(l)}}(x^{(l)}). \tag{6}$$

其中 $l \in \{1,2,\cdots,L\}$.

定理 2　交换操作不改变种群的方差，$D_t = D_0$. D_0 是初始种群方差.

证　由(6)式，定理 2 显然成立.

3.3　变异操作对种群方差的影响

由于对二进制编码和十进制编码，变异操作是不同的. 因此我们对二进制编码和十进制编码分别讨论变异操作对种群方差的影响.

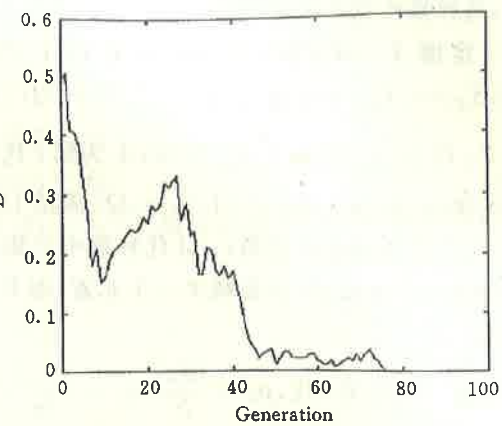


图 2　选择操作对种群方差的影响

对于十进制编码，变异操作可以描述为按变异概率 P_m 对每个个体的每一个基因位做如下改变：

$$x_{t+1}^{i(l)} = x_t^{i(l)} + N(0,\sigma^2), \tag{7}$$

$N(0,\sigma^2)$ 表示均值为 0 方差为 σ^2 的高斯分布随机变量.

定理 3　对十进制编码的种群，$P_m = 1$ 时，只进行变异操作，种群的方差满足下式

$$D_t = D_0 + t\sigma^2 I, \tag{8}$$

D_0 是初始种群的方差，I 为单位向量.

证　因为 x 和 $N(0,\sigma^2)$ 是相互独立的随机变量，由(7)式可得 $D_{t+1} = D_t + \sigma^2 I$，故定理 3 得证.

定理 3 说明了对于十进制编码的种群，变异操作使种群的方差线性单调增大.

对于二进制编码，变异操作可以描述为对每一个个体的每位按变异概率 P_m 做如下改变

$$x_t^{i(l)} = \sim x_t^{i(l)}. \tag{9}$$

其中，$l \in \{1,2,\cdots,L\}$，"～"表示二进制的求反操作.

定理 4　当 $P_m = 1$ 时，二进制编码的变异操作使种群的方差按如下变化

$$D_t = \begin{cases} D_0, & t \text{ 为偶数}, \\ D_1, & t \text{ 为奇数}. \end{cases} \tag{10}$$

其中 D_0 是初始种群的方差，D_1 是初始种群中所有个体的每一位都取反后形成种群的方差.

证　由(9)式可知经变异操作后，$t = 1$ 时种群中所有个体的每一位都取反，这时方差为 D_1，再经一次变异后，种群又变为初始种群. 这时方差 $D_3 = D_0$，种群的方差在 D_0 和 D_1 之间交替. 所以式(10)成立.

但当 $P_m < 1$ 时，情况就不那么简单了，图 3(a)是当 $P_m = 0.8$ 时种群方差的变化情况，图 3(b)是当 $P_m = 0.2$ 时种群方差的变化情况，图 3(c)是当 $P_m = 0.01$ 时种群方差的变化情况. 从图中可见，当 $P_m \neq 1$ 时，变异操作对种群方差的影响是很复杂的.

4　遗传算子对种群熵的影响

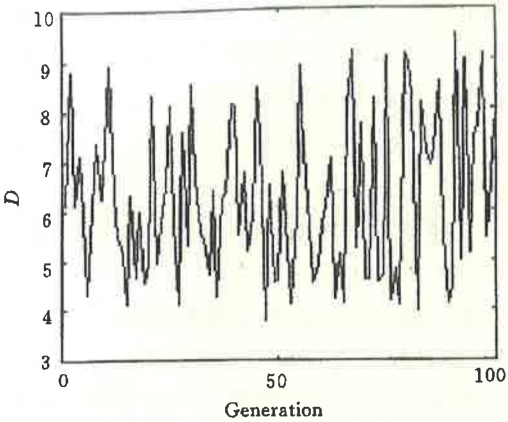
4.1　选择操作对种群熵的影响

熵反映了种群中不同类型个体的分布情况，而选择操作使种群收敛于最优个体，直观上

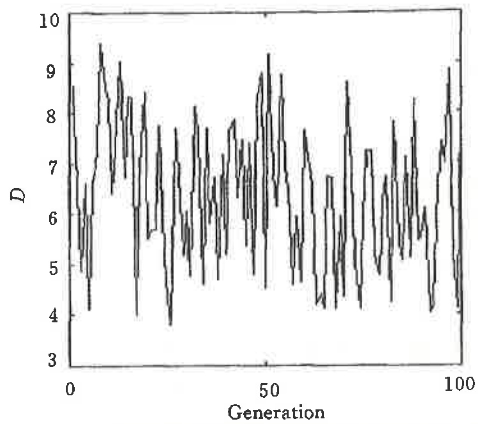
看，选择操作减少种群的熵.

定理 5 假定第 t 代种群中有 Q 个子集 $S_{t1},S_{t2},\cdots,S_{tQ}$，对任意 $p,q\in\{1,2,\cdots,Q\}$，满足 $S_{tp}\cap S_{tq}=\varnothing$ 和 $\bigcup_{q=1}^{Q}S_{tq}=A_t$，$A_t$ 为第 t 代种群的集合，且存在 $k,j\in\{1,2,\cdots,Q\}$ 满足 $|S_k|>|S_j|$. 若选择操作使第 $t+1$ 代种群中子集 S_k 中增加一个元素，S_j 子集减少一个元素，则 $E_{t+1}<E_t$.

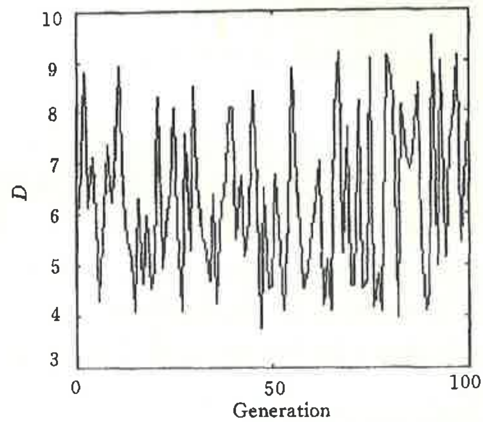
证 设在第 t 代，$p_k=\dfrac{|S_{tk}|}{N}$，$p_j=\dfrac{|S_{tj}|}{N}$，那么由(3)式得



(a) $P_m=0.8$



(b) $P_m=0.2$



(c) $P_m=0.01$

图 3 变异操作对二进制编码种群的方差的影响

$$E_{t+1}=E_t+p_k\log(p_k)+p_j\log(p_j)-(p_k+\frac{1}{N})\log(p_k+\frac{1}{N})-(p_j-\frac{1}{N})\log(p_j-\frac{1}{N}).$$

因为 $|S_k|>|S_j|$，故 $p_k>p_j$，容易证明 $E_{t+1}<E_t$.

定理 5 说明当种群中各子集的元素个数按其适应值排序后，选择操作使种群的熵单调下降，在此之前，熵可能上升或下降，如图 4 所示.

4.2 交换操作对种群熵的影响

定理 6 若种群中每个个体有 L 个基因位，且初始种群中有 Q 个子集：$S_{t1},S_{t2},\cdots,S_{tQ}$，对任意 $p,q\in\{1,2,\cdots,Q\}$，有 $S_{tp}\cap S_{tq}=\varnothing$，$\bigcup_{q=1}^{Q}S_{tq}=A_t$，$A_t$ 为第 t 代种群的集合，而且对任意 $m,n\in\{1,2,\cdots,Q\}$ 和 $l\in\{1,2,\cdots,L\}$，$x_t^i\in$

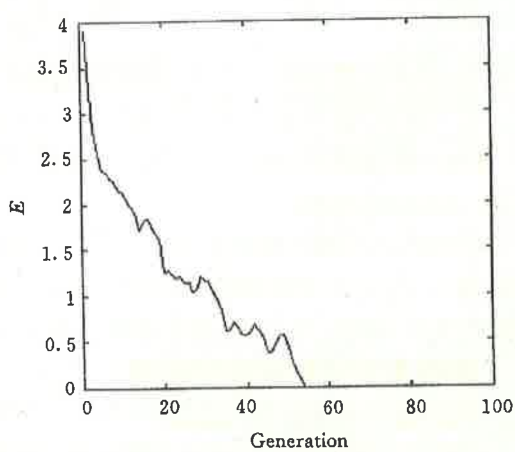


图 4 选择操作对种群熵的影响

$S_m, x_t^j \in S_n$，满足

$$x_t^i \neq x_t^j, \tag{11}$$

$$x_t^{i(l)} \neq x_t^{j(l)}. \tag{12}$$

则经过交换操作，当 $t \to \infty$ 时，种群中有 Q' 个不同类型的个体，

$$Q' = 2(L-1)C_Q^2 + Q. \tag{13}$$

若子集 S_k 中个体的 L 位基因是由 $l_1, l_2, \cdots, l_L \in \{1, 2, \cdots, Q\}$ 子集中的相应基因元素构成的，则子集 S_k 所对应的 $p_k = \frac{|S_k|}{N}$ 由下式可得

$$p_k = p_{l_1} p_{l_2}, \cdots, p_{l_L}. \tag{14}$$

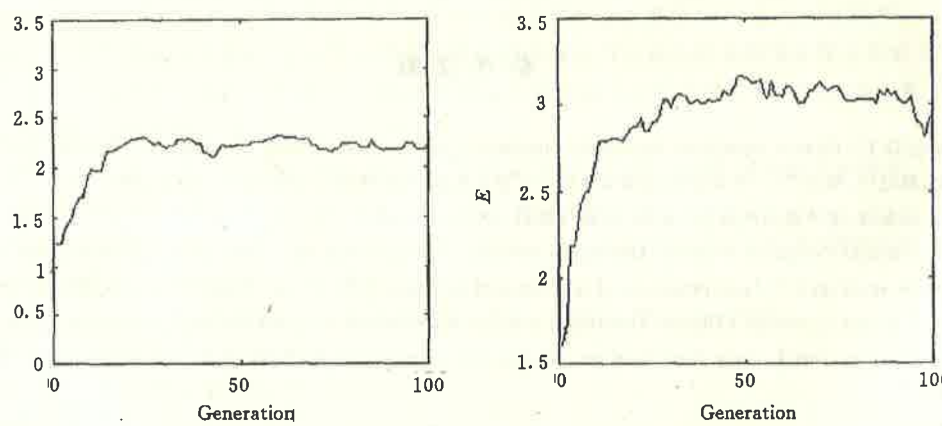
当 $t \to \infty$ 时，种群的熵 E 满足

$$\lim_{t\to\infty} E = -\sum_{k=1}^{Q'} p_k \log(p_k). \tag{15}$$

证　Q 个子集中任从两个子集中各取一个个体，交换的位置有 $(L-1)$ 个，由定理 6 的条件式(11) 和式(12) 可知，一定产生两个新个体 x', x''，且 $x', x'' \notin S_q, q \in \{1, 2, \cdots, Q\}$. 所以交换可产生 $2(L-1)C_Q^2$ 个新型个体，再加上原有的 Q 个类型，得出式(13). 由(4) 式可知，交换操作不改变每维坐标的基因元素，显然(14) 式成立. 由熵的定义容易得出(15) 式. 得证.

从定理 6 可以看出，交换操作不能保证种群的熵在 $t \to \infty$ 时达到最大，熵的变化与初始种群的分布有关，当初始种群的所有个体都相同，交换操作不改变种群的熵.

图 5(a)是初始种群中有 5 类不同的个体，$p_1 = p_2 = p_3 = p_4 = 0.1, p_5 = 0.6$ 时种群熵的变化情况. 图 5(b)是初始种群有 5 类不同的个体，$p_1 = p_2 = p_3 = p_4 = p_5 = 0.2$ 时种群熵的变化情况. 图 5 清楚地表明定理 6 的正确性.



(a)　$p_1 = p_2 = p_3 = p_4 = 0.1$，$p_5 = 0.6$　　(b)　$p_1 = p_2 = p_3 = p_4 = p_5 = 0.2$

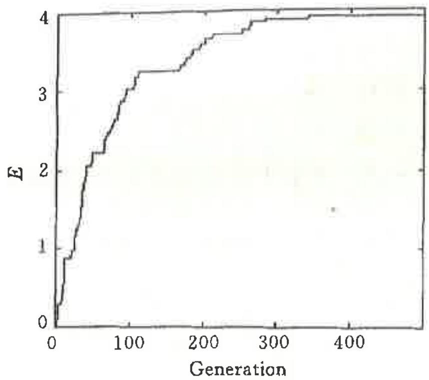
图 5　交换操作对种群熵的影响

用二进制编码时，定理 6 中的(11)和(12)式不易满足，在基因数目 L 相同的情况下，二进制编码产生的新个体类型数目比十进制少. 但这并不意味交换操作对二进制编码种群熵的提高能力小. 因为对一位十进制数，往往采用多位二进制表示，所以对同一问题，二进制编码的基因位总是高于十进制的基因位.

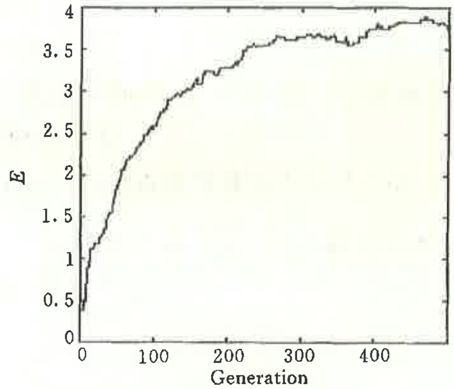
4.3　变异操作对种群熵的影响

图 6 是变异操作对种群熵的影响曲线. 图 6(a)是十进制编码的情况，图 6(b)是二进制编码的情况. 从图中可以看出，当 $t \to \infty$ 变异操作使种群熵达到最大值，变异概率越大，熵增加的

越快.遗憾的是我们目前还没有得出定量的结论.



(a)十进制编码



(b)二进制编码

图 6 变异操作对种群熵的影响

5 结 论

通过上面的分析,我们可以得出以下结论:

1) 种群的多样性可以由方差和熵两个量来全面衡量,方差体现了种群在搜索空间中的分布离散程度;熵体现了种群集合中不同类型个体的分布情况.

2) 选择操作使种群的方差和熵朝着减小的方向进化,减小了种群的多样性.

3) 交换操作不改变种群的方差,可提高种群的熵,但并不保证使熵达到最大值,交换操作对熵的影响与种群的初始分布有关.

4) 变异操作对于十进制编码,使种群的方差单调线性增大;对二进制编码不能保证方差增大,当 $t \to \infty$ 时,变异操作使种群熵达到最大值.变异操作可显著地提高种群的多样性.

参考文献

1 Goldberg,D. E.. Genetic algorithms in search,optimization,and machine learning. Reading,MA:Addison-Wesley,1989

2 张晓缋,戴冠中,徐乃平.一种新的优化搜索算法——遗传算法.控制理论与应用,1995,12(3):265—273

3 陈根社,陈新海.遗传算法的研究与进展.信息与控制,1994,23(4):215—222

4 Qi Xiaofeng and Francesco Palmieri. Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space,part Ⅰ:basic properties of selection and mutation. IEEE Trans. Neural Networks,1994,5(1):102—119

5 Qi Xiaofeng and Francesco Palmieri. Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space,part Ⅱ:analysis of the diversification role of crossover. IEEE Trans. Neural Networks,1994,5(1):120—129

Study on Diversity of Population in Genetic Algorithms

ZHANG Xiaohui, DAI Guanzhong and XU Naiping

(Department of Automatic Control,Northwestern Polytechnical University · Xi'an,710072,PRC)

Abstract: In this paper,variance and entropy are proposed as measures of diversity of pupulation in genetic algorithms. The influence which selection,crossover,and mutation act upon variance and entropy is analyzed. In addition,the different influence between binary and decimal encoding mechanism is also shown.

Key words: genetic algorithms; selection; crossover; mutation; diversity of population

本文作者简介

张晓缋　女. 1968 年生. 分别于 1990，1993 年获西北工业大学计算机应用专业学士学位和硕士学位，现为西北工业大学自动控制系博士研究生. 主要研究方向有进化计算，科学计算可视化，虚拟现实.

戴冠中　1937 年生. 现为西北工业大学校长，教授，博士生导师. 主要研究领域有大系统估计与控制理论，智能控制，控制系统中的并行处理理论、算法与并行计算机.

徐乃平　1935 年生. 现为西北工业大学自动控制系教授，博士生导师. 主要研究方向为并行处理与并行计算机，非线性科学，人工生命，虚拟现实.

《智能控制——基础及应用》出版

蔡自兴教授的专著《智能控制——基础及应用》已由国防工业出版社出版. 该书通过中国国防出版基金评审委员会评审后，由该基金资助出版.

本书涉及智能控制的基本概念、原理、方法、技术及其应用，着重阐述智能控制的五个典型系统的原理、方法及应用. 全书共八章. 第一章简述智能控制的产生背景、起源与发展，讨论智能控制的定义、特点和智能控制的四元交集结构理论，揭示各相关学科间的内在联系. 第二章至第六章逐章讨论了递阶控制系统、专家控制系统、模糊控制系统、神经控制系统和学习控制系统的作用机理、类型结构、设计要求、控制特性和应用示例. 对于不同的系统，每章研究的侧重点也有所区别. 第七章概括智能控制的主要应用领域，指出智能控制应用研究中存在的一些问题. 最后一章，即第八章，展望智能控制的发展方向及其与相关技术的关系.

本书相当大一部分内容是作者及其研究组成员们近年来的研究成果；加上其它新内容，本书校好地反映出国内外智能控制研究和应用最新进展，与蔡自兴教授的另一著作《智能控制》(电子工业出版社，1990，全国高校统编教材，中国第三届优秀教材一等奖)相比，有 85% 以上内容得以更新，水平也有明显提高.

本书是以专著形式编写出版的，它不仅可供从事智能控制研究与应用的自动控制科技工作者学习参考，而且特别适合作为高等院校自动控制、电子工程、计算机应用和机电自动化等类学科(专业)高年级学生和研究生的《智能控制》课程教材. 全书 35 万字，定价 20 元.

联系单位：100036 北京市海淀区紫竹院南路 23 号 国防工业出版社发行部，或计算机编辑室.

电　　话：(010)68469922

本书的姊妹篇英文版《Intelligent Control: Principles, Techniques and Applications》(新加坡-美国新泽西州，World Scientific 出版社)已经发行.

Published by World Scientific Publishing Co. Ptd., P. O. Box 128, Farrer Road, Singapore 912805

450pp (approx).　　Pub. date: November 1997

981-02-2564-4　　US $64, £45, S $92

Book Code: EeRa-B3028

Main Subject Classification: Engineening and Electronics

Bood B+17/8/95

Editor: YK

BL/d

(刘健勤　王　晶)