第10卷 第4期　　自 动 化 学 报　　Vol. 10, No. 4
1984年10月　　ACTA AUTOMATICA SINICA　　Oct., 1984

综述与评论

学习自动机概述

窦如静　何成武
(华东计算技术研究所)

摘　要

本文概述了学习自动机方面的已有成果，着重叙述学习自动机的基本原理和近期的研究进展，最后介绍了学习自动机的应用.

一、引　言

Цетлин 于 1961 年提出确定自动机在与随机环境相互作用的过程中可以具有某种学习性质；1963 年 Варшавский 和 Воронцова 提出具有可变结构的随机自动机与随机环境相互作用的模型. 前一种结构有时称为固定结构的学习自动机，后一种结构则通称为学习自动机. 本文的论述对象主要是后一种结构.

在确定型控制论中，过程控制之前已有关于过程特征的完全知识. 在随机控制论中考虑了过程中存在的不确定性，而这种不确定性的概率特征是已知的. 但是，实际上不确定性常常是高阶的，而且概率特征也可能是完全未知的. 于是必须在过程进展中对它进行观察，以获得关于它的进一步知识. 可以把这种方式看成学习. 从纯数学的观点看，学习系统的目标是优化一种不明显知道的函数，例如概率分布未知的某个随机函数的数学期望. 其基本思想可用如下例子说明. 如教师对学生提出问题，并提供有限数目的备选答案. 学生选择其中一个答案，教师对此选择的反应方式是二元的，即指出对或错. 但教师是“概率的”，对于学生的任何选择，两种反应的概率均不为零. 唯一已知的是，对于“正确的答案”，教师指出“错”的概率最小. 而到底哪个答案是“正确的”，学生事先并不知道，必须找出一种算法处理从教师处得到的信息，并形成回答序列，使得教师指出“错”的概率最小，即“学习”到正确的答案.

学习自动机模型曾在数学心理学中用于解释人和动物的学习行为，后来作为参数优化、假设检验及对策论的工具被陆续应用到数理统计、自动控制、通讯网络及经济系统等领域的研究. 限于篇幅，本文重点介绍近期的研究进展，凡未列出的文献请参见文献 [1] 及其它有关文献的附录.

二、数学模型

图 1 展示了学习自动机与随机环境(简称环境，也称介质)组成的学习系统模型. 学

本文于 1983 年 12 月 16 日收到.

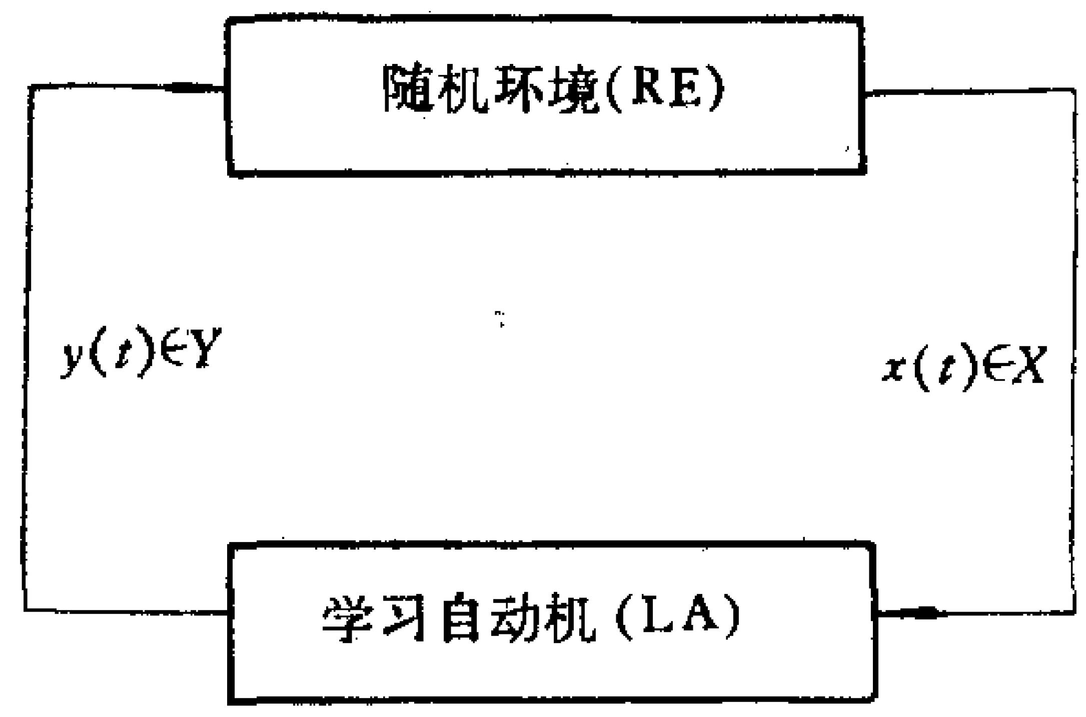

图 1　学习系统模型

习自动机(LA)由六元组 (X, Q, Y, $\boldsymbol{\pi}$, R, G) 表示. 其中$X=\{x_1, \cdots, x_l\}$ 为输入集；$Q=\{q_1, \cdots, q_n\}$ 为状态集；$Y=\{y_1, \cdots, y_m\}$ 为输出集；G是输出函数. 不失一般性，可设G为从Q到Y的一对一确定映射，从而有 $m=n$，此时"状态"与"输出"可看成是同义的，并泛称为"行动"；$\boldsymbol{\pi}$ 是如下的状态概率分布矢量:

$$\boldsymbol{\pi}(t)=(\pi_1(t),\ \pi_2(t),\ \cdots,\ \pi_n(t))^\tau,$$

其中 τ 表示转置，$\pi_i(t)$ 表示在时刻 t 自动机处于状态 q_i 的概率；R是学习算法，也称更新方略，用下式表示:

$$\boldsymbol{\pi}(t+1)=T(\boldsymbol{\pi}(t),\quad y(t),\ x(t)). \tag{1}$$

式中T是一个算子. 自动机工作于离散时刻 $t=0$, 1, 2, $\cdots$，时刻 t 的输出 $y(t)$ 作为环境的输入，环境的响应 $x(t)$ 作为自动机的输入. 环境(RE)由三元组(Y, X, C)描述，Y和X分别是其输入集和输出集(响应集). 当X为二元集(一般用{0, 1}表示)时，环境称为P型的；当X为多于二元的有限集时就得到Q型环境；当X取连续值(一般用区间(0,1)表示)时就得到S型环境. Q型环境适用于对策论研究，S型环境适用于许多实际的自动控制领域. P型环境是最简单的，其中很多研究成果都可推广到Q型和S型环境. C为$Y*X$到区间(0, 1)上的映射矩阵，其值是未知的. 如果C的取值不随时间而变，则此环境称为平稳的，否则为非平稳的. 本文一般讨论P型平稳环境，其中 $x(t)=0$ 称为环境对 LA 的奖励；$x(t)=1$ 称为对 LA 的惩罚. C 可简记为 $\mathbf{c}=(C_1,\ \cdots, C_m)$，其中 C_i 表示当 $y(t)=y_i$ 时，$x(t)=1$ 的概率，故可称 $\mathbf{c}$ 为惩罚概率矢量.

如果自动机输出各个 y_i 的概率是相等的(等权输出)，则可以从环境得到如下的平均惩罚:

$$M_0=\frac{1}{m}\sum_{i=1}^{m}c_i.$$

就一般情况而言，自动机在时刻 t 收到的平均惩罚为

$$M(t)=\sum_{i=1}^{m}\pi_i(t)c_i.$$

$M(t)$称为自动机的"惩罚函数"，为随机变量，其均值用 $E[M(t)]$ 表示. 以下几个定义描述了自动机的不同性能.

定义 1. 如果 $\lim\limits_{t\to\infty}E[M(t)]<M_0$，则 LA 称为"有利的" (Expedient).

定义 2. 如果 $\lim\limits_{t\to\infty}E[M(t)]=\min\limits_i\{c_i\}$，则 LA 称为"优化的"(Optimal).

定义 3. 如果对于任意的 $\varepsilon>0$，通过适当选择学习算法的参数，可有

$$\lim_{t\to\infty}E[M(t)]<\min_i\{c_i\}+\varepsilon,$$

则 $\angle A$ 称为"ε 优化的"(ε-optimal).

定义 4. 如果对于所有的 t 和 $\pi_k(t)\in(0,1)$以及各种可能的 $c_i(k,i=1,\cdots,m)$，有

$$E[M(t+1)|\boldsymbol{\pi}(t)]<M(t),$$

则 $\angle A$ 称为"绝对有利的"(Absolutely Expedient).

显然，人们希望得到优化的 $\angle A$，因为它能导致惩罚函数的最小期望值. 遗憾的是迄今为止所得到的算法(除个别例外)都不是优化的. 但是 ε 优化的或绝对有利的自动机具有令人满意的实用结果，因而近年来有关学习算法的研究主要致力于这两类性质的获得.

三、学习算法

在(1)式中，如果算子 T 是 $\boldsymbol{\pi}(t)$ 各分量的线性函数，则称该算法为线性的，否则为非线性的. 一般学习算法的基本思想是很简单的：如果时刻 t 的行动 y_i 导致环境的奖励，则增加下一时刻该行动的概率 $\pi_i(t+1)$ (并相应减小 $\boldsymbol{\pi}(t+1)$ 的其余分量)；否则就减少 $\pi_i(t+1)$ (并相应增加 $\boldsymbol{\pi}(t+1)$ 的其余分量). 称这些增、减为对 $\pi_i(t)$ 的奖、惩. 也有的算法在得到环境的奖励(或惩罚)后仍保持 $\pi_i(t+1)$ 等于 $\pi_i(t)$，这种做法称为"不作用".

将(1)式的算子稍加具体化，得到学习算法的如下表示：当 $y(t)=y_i$ 时，令

$$\pi_j(t+1)=\begin{cases}\pi_j(t)-f_j(\boldsymbol{\pi}(t)), & \text{当 } x(t)=0 \text{ 时},(j\neq i)\\ \pi_j(t)+g_j(\boldsymbol{\pi}(t)), & \text{当 } x(t)=1 \text{ 时};\end{cases} \tag{2a}$$

$$\pi_i(t+1)=\begin{cases}\pi_i(t)+\sum\limits_{j\neq i}f_j(\boldsymbol{\pi}(t)), & \text{当 } x(t)=0 \text{ 时},\\ \pi_i(t)-\sum\limits_{j\neq i}g_j(\boldsymbol{\pi}(t)), & \text{当 } x(t)=1 \text{ 时}.\end{cases} \tag{2b}$$

其中要求非负连续函数 f_j 和 g_j 必须保证自动机不致死锁于吸收状态. 为节省篇幅，常略去算法表示式中的自变量 t 或 $\boldsymbol{\pi}(t)$. 设

$$f_j=a\pi_j,\quad g_j\equiv 0,$$

就得到线性奖——不作用型(L_{R-I})算法. 它是应用最广的一类 ε 优化算法. L_{R-I} 算法在 Q 型和 S 型环境下的相应扩展称为 SL_{R-I} 算法，它也是 ε 优化的.

Lakshmivarahan 等人提出了形如(2)式的一般非线性算法成为绝对有利算法的充要条件. 后来这个充要条件扩展到 Q 型和 S 型环境下的算法. 文献[2]证明了绝对有利性隐含了 ε 优化性，从而得到了设计 ε 优化算法的充分条件.

文献[3]将一般的算法扩展到"更新函数与选择第几个行动有关"(即 f_j 和 g_j 可写成 f_j^i 和 g_j^i)，并导出扩展后的一般算法成为绝对有利的充要条件. 文献[4]提出强绝对有利性的概念(限于篇幅，介绍从略)，导出上述扩展算法满足强绝对有利性的充要条件，并

证明强绝对有利性也隐含 ε 优化.

以上是基于"平均惩罚"即 $E[M(t)]$ 的演变而考虑自动机的行为特征．更深入一步，可以直接考虑状态概率分布矢量 $\boldsymbol{\pi}(t)$ 本身的渐近性质．不同的学习算法导致 $\boldsymbol{\pi}(t)$ 的两种收敛情况：一类算法产生的马尔可夫过程 $\{\boldsymbol{\pi}(t)\}_{t\geqslant 0}$ 是遍历的，没有吸收状态，这类算法导致 $\boldsymbol{\pi}(t)$ 依分布收敛；另一类算法产生的马尔可夫过程具有不止一个吸收状态，$\boldsymbol{\pi}(t)$ 以概率 1 收敛于吸收状态集．因为至少有两个吸收态，而其中只有一个是希望状态(它对应于最小惩罚概率)，所以只能说 $\boldsymbol{\pi}(t)$ 以正概率收敛于希望状态．但对于这个概率，迄今为止只得到取值范围的一些界限.

关于收敛速度，目前所得到的结果更为有限．一般说，收敛的速度和精度是有矛盾的，从速度和精度两方面对不同算法进行比较显得更困难.

四、更复杂的学习问题

1. 自动机对策

自动机对策的模型如图 2 所示，其中有 N 个局中人，即自动机 $A_1, A_2, \cdots, A_N$，随机环境作为仲裁．在时刻 t，各自动机选用的策略(即行动)组成该时刻的局 (Play)． 环境对各自动机的反作用称为各自动机的支付 (payoff)．各自动机独立地更新其混合策略 (即状态概率分布矢量)． 不断更新所产生的局的序列组成对策.

在 Von Neumann 的经典对策论中，已假定向局中人提供了关于对策的完全信息，如局中人数目、可能策略及支付矩阵等．在自动机对策中，这些先验信息是得不到的． 自动机必须在线地(On Line)选择其策略，其支付函数是具有未知分布的随机变量．因此，自动机对策是"具有不完全信息"的对策.

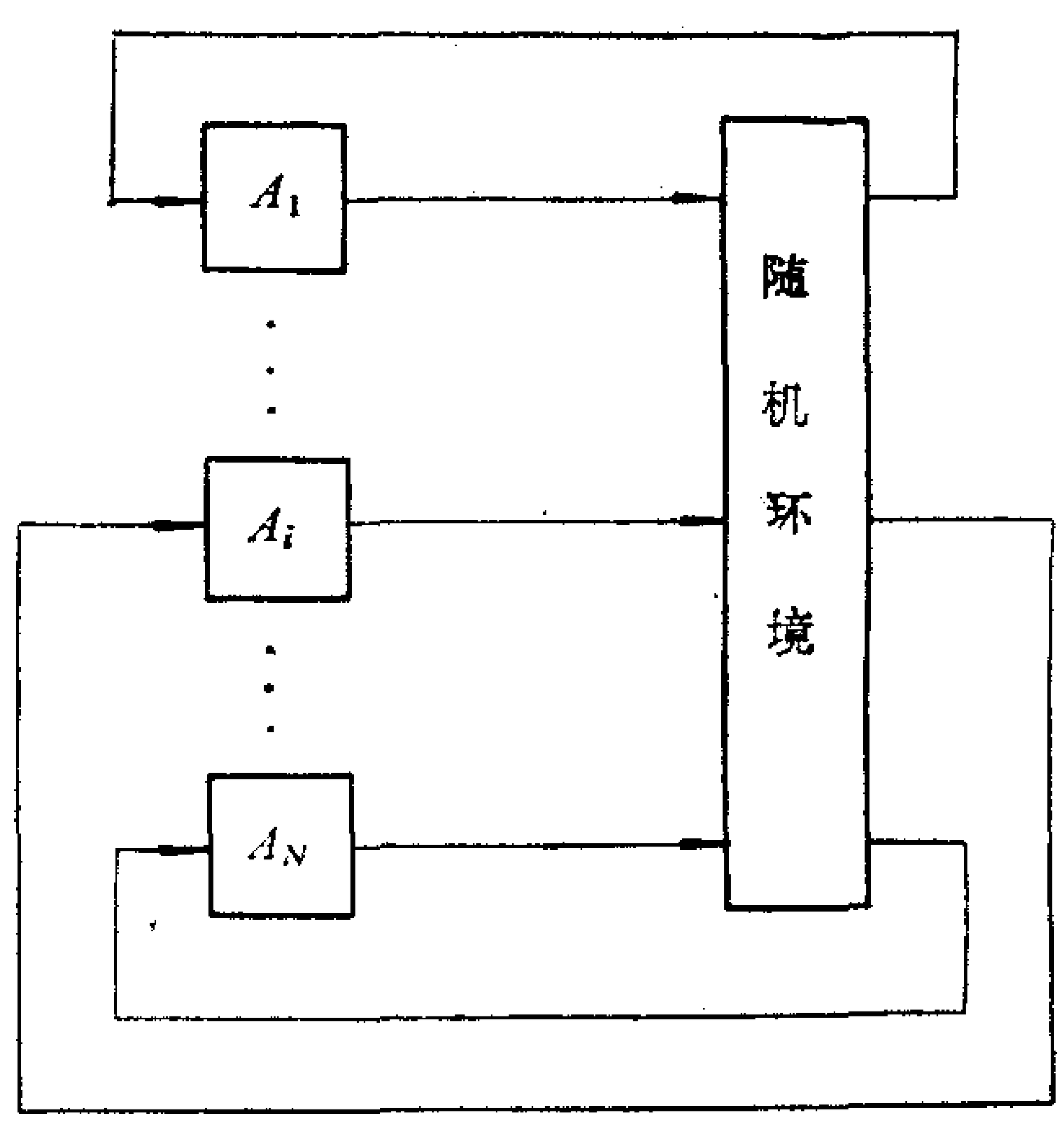

图 2　自动机对策

自动机对策的概念最初是由 Крылов 和 Цетлин 针对确定自动机提出来的． 后来这一结果扩展到学习自动机．在前人一系列研究工作的基础上，文献 [5] 详细总结了二人零和对策，证明在有鞍点对策矩阵或无鞍点对策矩阵的情况下，局中人按特定的学习算法更新其混合策略，最终将导致 Von Neumann 值．文献 [5] 还讨论了一种特殊的非零和对策，即所有局中人具有相同的支付函数，这类问题称为 Team 决策论，它是经典的贝叶斯统计决策论到一组相互依存的决策者的推广．决策的目标是寻求适当的学习算法，使各局中人的渐近期望支付最大.

自动机对策模型是研究系统理论的有力工具，也是学习自动机领域中当前最活跃的课题．由于本课题的内在复杂性，迄今只取得部分成果，还有很多问题有待于进行更深入的研究.

2. 非平稳随机环境下的学习

在多自动机对策中，各自动机的支付不仅取决于环境的概率特征，也取决于各自动机所选纯策略的组合. 由于纯策略的组合是不断更新的，所以对任何单个自动机而言，等价于工作在非平稳环境之中. 例如通讯网络的路由选定及马尔可夫链的控制等问题，其中自动机都工作于非平稳环境.

由于非平稳环境的研究十分困难，所以常常施加某些限制条件，以导出一些有用的结果. 例如，设环境的惩罚概率矢量按时间周期地变化，而周期值未知，则可使用所谓"两级自动机系统". 第一级估计未知周期，第二级假设此估计为真而在一个循环内稳定地工作于该环境. 将这个循环内环境的平均输出作为第一级的输入，以确定周期的下一个估计.

文献[6]提出另一类非平稳环境，其惩罚概率矢量记为$(c_1(t, \omega), \cdots, c_m(t, \omega))$，其中 $\omega \in \Omega$，Ω 是概率测度空间的支集 (Suporting Set). 该文证明了如果对某个行动 y_α 及某些 $\delta > 0$ 以及所有的 t 和 $\omega \in \Omega$，有

$$c_\alpha(t, \omega) + \delta < c_{k1}(t, \omega), \cdots, c_{k(m-1)}(t, \omega),$$

则使用绝对有利算法可保证自动机的 ε 优化性能.

文献[7]定义了"可分离的" (Separable) 非平稳随机环境，并将平稳环境下"绝对有利性"的概念推广到这类非平稳环境. 通过引入概率的李雅普诺夫函数，对学习自动机收敛于优化状态的概率及收敛的概率稳定性作了定量评价.

3. 多教师下的学习

文献[8]推广了环境的概念，把自动机置于几个教师的作用下，自动机的目标是询问教师并确定潜在环境最佳的具体行为模式. 文献[9]和[10]考虑了变结构随机自动机在多教师环境中的学习策略，通过引入"平均加权奖"推广了单教师环境下的绝对有利和 ε 优化的概念. 作为以往算法的扩展形式，文中提出称为 GAE 的非线性算法和称为 GL_{R-I} 的线性算法的一般类，证明了 GAE 和 GL_{R-I} 在一般的多教师环境中是绝对有利的或 ε 优化的，模拟结果也验证了 GAE 和 GL_{R-I} 算法的有效性.

4. 学习自动机的递阶系统

前面提到使用两级自动机系统，解决周期未知的环境下的学习问题，人们也常用多级自动机解决高维数的参数优化中收敛速度缓慢的问题. 文献[11]介绍了保证 ε 优化的递阶系统，其学习算法是绝对有利算法的简单修改，递阶系统各级的算法参数只取决于前一级的参数和行动概率，为最小化每个周期的更新数目，系统中各自动机只需二到三个行动.

五、应用研究

1. 参数优化

自适应控制、模式识别、滤波及系统辨识等方面提出的许多问题，在适当的假设条件下可以看成参数优化问题，应用学习自动机可有效地解决这些问题，特别是在先验信息很少的有扰条件下.

给定能一个系统，只能得到性函数

$$I(\alpha)=E\{g(\alpha,\omega)\}$$

的有扰测度 $g(\alpha,\omega)$，这里 α 是 m 维参数矢量；ω 是测度干扰．参数优化问题是确定优化的参数矢量 α_{opt}，以使系统性能函数取极值．因为缺少系统结构及性能函数的充分信息，所以不可能得到问题的解析解．可以用学习自动机作为自适应控制器，其行动对应于各别的 α 值，自动机基于性能函数的测度而更新各参数值的概率．自动机方法的吸引力在于它不致死锁在局部优化点，而是导致全局性优化．

在使用 P 型算法时，为了解决性能测度(实际上常常为连续变量)与环境的二值响应相关联的问题，傅京孙定义了惩罚强度模型．该模型中环境响应可取 [0，1] 区间中的任何值．在应用 L_{R-I} 算法找出多峰表面的全局最小值、并证明自动机以最高概率选择全局最小值方面也取得了成功．使用 L_{R-I} 等 ε 优化算法可成功地处理迭加了高方差噪声的相对平坦的性能表面．

如果性能测度的界已知，则可简单地把它们归一化为（0，1）区间，然后使用 S 型算法．如果这些界是未知的，则可以使用这些界的当前估计以便归一化．实验表明这样仍能达到 ε 优化收敛．文献[12]从理论角度总结了学习自动机用于参数优化问题的研究成果．

2. 通信网络的路由选择

通信网络由节点和链路组成，节点代表交换中心，链路即为相邻节点间的通信路径．根据每个节点提出的要求，建立各对节点之间的通信路径，称为“路由选择”，简称“编路”(Routing)．如果通信网络的状态是平稳的．则使用常规的编路技术是有效的．但是考虑到话务量和交换模式的非平稳情况，特别是部件失效、自然灾害及政治事件等干扰，则引入自适应编路技术是很自然的．假设在网络的某个节点使用一个学习自动机，由该节点出发并到达指定目标的不同“路由”构成自动机的“行动”，一旦有关的“呼叫”(call）出现，就根据行动概率矢量选定路由．如果这个呼叫成功地通过网络发送到目标处，则此次所用的路由的概率应予提高，即给予“奖励”；否则，将降低该路由的概率，即给予“惩罚”．文献[13]总结了电话交换系统使用学习自动机编路的研究进展，介绍了在各节点上具有缓慢学习算法的网络模型．在这个模型下，自动机工作于非平稳环境，其行为模式也可用自动机对策的方法加以说明．该模型的一个显著优点是可以预期将显著改进阻塞情况的那些节点引入学习自动机．

自动机编路的方法是易实现的，它主要涉及软件的修改，从经济观点考虑，这是很吸引人的．文献[13]也指出了所论网络模型的改进方向和其他有待于深入研究的课题．

3. 马尔可夫链的控制

有一类广泛的控制对象可用马尔可夫链描述，这是一类特殊的非平稳环境．作为模型的马尔可夫链有 k 个状态($2\leqslant k<\infty$)，状态转移矩阵一般是控制行动的函数．在给定状态下，控制对象类似平稳随机环境，即对各个控制行动有固定的奖励（或惩罚)概率．而在不同状态下，这些概率特征一般是不同的．用学习自动机做控制器的模型，用学习算法更新其控制矢量 $y=(y^1,\cdots,y^k)$，其中 y^i 是已知环境处于状态 i 时自动机采取的行动．于是控制目标是寻求如下的一步优化控制矢量

$$y_*=(y_*^1,\cdots,y_*^k),$$

其中 $y^i_{*}=y_{i*}$ 是当环境处于状态 i 时对应于最大奖励概率的控制行动.

假设环境的状态转移矩阵及各状态下的奖励概率特征是未知的，环境当前所处的状态是已知的，文献[5]证明只要适当选择算法参数，可以在长运行中以尽可能逼近 1 的概率到达一步优化控制矢量. 文献[5]还讨论了环境的奖励概率特征未知、转移矩阵已知、但环境的状态延迟一步才可得知时的控制方法，此时将涉及到利用前一时刻环境状态转移情况作出后一时刻环境所处状态的预测.

Цетлин 等人研究过的"复式环境"是上述环境的特例，其中状态转移矩阵是恒定的，与控制行动无关. 显然，周期环境也是这类环境的特例，其状态转移矩阵不仅与控制行动无关，而且各行恰有一个元素为 1，其余元素都为 0，但是一般说来，周期(即环境的状态数目)是未知的，这隐含环境所处的状态是不可知的. 关于使用学习自动机控制马尔可夫链的问题有过较广泛的研究，有关结果的综述，请参看文献[5].

4. 经济系统

文献[14]把自动机对策的研究成果用于竞争性经济的市场价格形成及系统工作中的一维资源优化分配. 前者指卖主根据他对某项商品的提供能力及用户在一定价格下的平均需求，而不断调整商品价格，以获取最大利润；后者指如何将数量为 R 的资源分配给 N 个用户，以使总产量达最大值. 实际上，该文所述问题是 Малишевский 工作的推广. 但这个推广在理论上是不够严密的，有待于进一步修改和补充. 尽管如此，这篇文章仍展示了学习自动机模型在研究管理与运筹方面的潜力.

5. 工业机器人

文献[15]报道了使用学习自动机构成力敏 (Forcesensing) 机器人，执行高精度组装操作的控制算法. 这种算法把输入和传感器信号翻译成正确的控制命令. 由于常常有大量参数是不可知的，因而所有常规的预编程序很容易变得冗长乏味. 而且对于较高级的任务，对过程反馈信号的测度的解释可能是很复杂的，为处理这些及其它有关问题，未来一代的工业机器人将需要某种形式的自发智能. 由于"学习"意味着有可能考虑过去的经验，因而就成了高级控制技术的研究课题. 在文献[15]中，学习自动机的可变结构已不仅是指其状态概率矢量 $\pi(t)$ 的随机性，而且扩展到逐渐优化输入变量的分辨率(量化级)，并在某种程度上扩展到选择自动机的内部状态集合. 这是解决参数优化中高维数问题的一个新的成功方法.

6. 图的划分

令一无向图 $G(P,U)$ 具有加权的顶点和边，没有循环. 要求划分 $G(P,U)$ 为若干子图，使得各子图的顶点的总权不超过某数 K，而连接这些子图的总权最小. 目前所有已知解决类似问题的方法要求指数复杂度的计算时间，还不知道是否有使计算时间取多项式复杂度的算法.

对于许多实际问题，最小图划分是很适用的理论模式. 例如，具有最小页间互连的虚存页面组织问题；数据分布网的分区问题；多机系统中程序互访最小化的程序分配问题，印刷电路板上电路模块之间互连最小化的布局问题，等等，都可归结为本问题. 文献[16]提出解决这个问题的一组自动机模型，模型的本质是通过具有可变结构的概率自动机的集合解决划分问题，将这些自动机标识为图的元素，并使用统计建模方法于自适应模型.

六，结　论

学习自动机模型自问世以来已有二十多年历史．在几乎整个六十年代，苏联学者就有关固定结构学习自动机的问题进行了广泛地研究．后来许多西方学者把上述研究推广到具有可变结构的学习自动机，并形成了学习自动机研究的主流．这是因为固定结构的学习自动机要达到优化目标，理论上需要无限多记忆(内部状态)，而变结构学习自动机只要有限状态就可达到或接近这个目标．学习自动机领域具有众多的研究课题和广泛的应用对象，除前面叙述的典型应用外，文献[17]还报道了学习自动机在发展长途电信系统工程中的应用．迄今为止学习自动机的研究已取得了大量成果．但由于课题的内在复杂性，很多问题还有待于进行开拓性研究．此外学习自动机的理论涉及较多、较新的数学概念，而且资料分散、叙述多不统一等等对其推广应用有一定影响．从国外刊物上看，近年来这些情况已有所改善．

参　考　文　献

[1] Narendra, K. S. et al., Learning Automata—A Survey, *IEEE Trans. Syst.*, Man, Cyber., SMC-4 (1974), 323—334.

[2] Lakshmivarahan, S. et al., Bounds on the Convergence Probabilities of Learning Automata, SMC-6 (1976), 756—763.

[3] Hirotomo ASO et al., Absolute Expediency of Learning Automata, *Information Sciences*, **17** (1979), 91—112.

[4] Meybodi, M. R. et al., ε-optimality of a General Class of Learning Algorithms, *Information Sciences*, **28** (1982), 1—20.

[5] Lakshmivarahan, S., Learning Algorithms: Theory and Applications, Springer-Verlag (1981).

[6] Baba, N. et al., On the Learning Behaviro of Stochastic Automata Under a Nonstationary Random Environment, SMC-**5** (1975), 273—275.

[7] Yoshihito Toyama et al., On Learning Automata in Nonstationary Random Environments, *Systems. Computers. Controls*, **8**(1977), 66—73.

[8] Koditschk, D. E. et al., Fixed Structure Automata in a Multi-Teacher Environment, SMC-7 (1977), 616—624.

[9] BABA, N. The Absolutely Expedient Nonlinear Reinforcement Schemes under the Unknown Multiteacher Environment, SMC-**13** (1983), 100—108.

[10] Baba, N., On the Learning Behavior of VSSA in General N-Teacher Environment, SMC-**13** (1983), 224—231.

[11] Thathachar, M. A. L. et al., A Hierarchical System of Learning Automata, SMC-**11** (1981), 236—241.

[12] BABA, N., Theoretical Considerations of the Parameter Self-optimization by Stochastic Automata, *I. J. Control*, **27**(1978), 271—276.

[13] Narendra, K. S. et al., The Use of Learning Algorithms in Telephone Traffic Routing—A methodology, *Automatica*, **19** (1983), 495—502.

[14] El-Fattah, Y. M., Stochastic Automata Modeling of Certain Problems of Collective Behavior, SMC-**10**(1980), 304—314.

[15] Simons, J. et al., A Self-Learning Automata with Variable Resolution for High Precision Assembly by Industrial Robots, *IEEE Trans. Automatic Control*, AC-27 (1982), 1109—1113.

[16] Vagin, V. N. et al., On Automata Models of Graph Partitioning, *Engineering Cybernetics* (Soviet Journal) **17**(1979), 44—52.

[17] Probert, D. E., Aspects of Evolving Systems in Telecommunications Modelling, Proceedings of international conference on cybernetics and society (1980), 52—458.

A SURVEY OF LEARNING AUTOMATA

Dou Rujing He Chengwu

(East-China Research Institute of Computer Technology)

Abstract

A survey of the available results in the area of learning automata has been attempted in this paper. Attention has been focused on the principle of learning automata and recent advances in its investigation. Applications of learning automata are also discussed.