Lecture Notes on Stochastic Learning Automata
Lecture #1

Aly El-Osery

1 Learning Control

– In the design of an optimal control system if all the information about the controlled plant is deterministic and known, then the controller can be designed using deterministic optimization methods.

– If all or part of the a prior knowledge of the plant can be described probabilistically, then stochastic design techniques are used.

– If the a prior knowledge of the plant required is not known, as is the case in many practical problems, then learning control has to be utilized.

– Since not all the required information is available, and hence, the controller designed should be capable of estimating the unknown information. In this case as the estimated information approaches the true information as time progresses, then the controller gradually approaches the optimal one.

Definition 1 (Learning Control)
A learning controller is one that learns the information during the operation and the learned information is, in turn, used as an experience for future decisions or controls.

– Depending upon whether or not an external supervision (in the form of a teacher) is required, the process of learning can be classified into:

- Learning with external supervision (or training or supervised or off-line learning).
- Learning without external supervision (or nonsupervised or on-line learning).

- In the learning process with external supervision, the desired answer (for example, the desired output of the system or the desired optimal control action) is usually considered exactly known.
- Directed by the known answer (given by external teacher, say), the controller modifies its control strategy or control parameters to improve the system's performance.
- On the other hand, in learning processes without external supervision, the desired answer is not exactly known.

2 Reinforcement Learning

- Mutually exclusive and exhaustive classes of responses $w_1, \ldots, w_m$ are generally considered.
- Let P_i be the probability of occurrence of the ith class or responses.
- We consider the performance change being expressed by the change or *reinforcement* of the set of responses probabilities $\{P_i\}$.
- Mathematically, the reinforcement of $\{P_i\}$ can be described as the following relationship:

  $$P_i(n+1) = \alpha P_i(n) + (1-\alpha)\lambda_i(n), \quad n = 0, 1, 2, \ldots \tag{1}$$

  where $P_i(n)$ is the probability of the occurrence of w_i at instant n when the input X is observed

  $$0 < \alpha < 1, \quad 0 \leq \lambda_i(n) \leq 1,$$

  and

  $$\sum_{i=1}^{m} \lambda_i(n) = 1$$

- Because of the relationship between $P_i(n+1)$ and $P_i(n)$ being linear, Eq. (1) is often called *linear reinforcement linear algorithm.*
- It can easily be shown that if $\lambda_i(n) = \lambda_i$, then

  $$P_i(n) = \alpha^n P_i(0) + (1-\alpha^n)\lambda_i, \tag{2}$$

  and

  $$\lim_{n\to\infty} P_i(n) = \lambda_i. \tag{3}$$

- It is noted that from Eq. (3), λ_i is the limiting probability of $P_i(n)$.

- λ_i should be, in general, related to the information or performance evaluated from the input X at instance n.

- In simple cases, λ_i may be 0 or 1 to indicate whether the performance of the system at instant n, due to the ith control action, is satisfactory or unsatisfactory.

3 Stochastic Automata as a Model of Learning Controllers

- The reinforcement learning control may be formulated mathematically by way of stochastic automata theory.

- An SLA is connected in a feedback loop (see Figure 1).

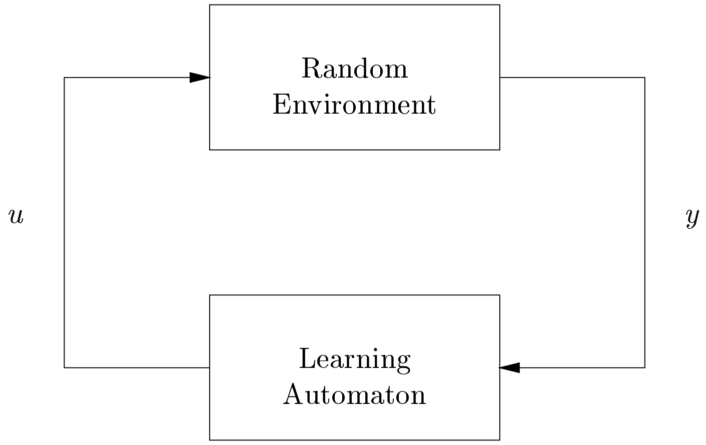

Figure 1: Automaton operating in random environment

- SLA needs no information about the plant to be controlled or the function to be optimized.

- At each step, the performance of the SLA through the environment is evaluated by either a penalty (or unsatisfactory performance) ($y = 1$), or a nonpenalty (or satisfactory performance) ($y = 0$).

- A stochastic automaton is a quintuple $\{Y, Q, U, F, G\}$ where

- If Y consists of only two elements 0 and 1, the environment is said to be *P-model.* When input into the SLA is a finite number values in the closed interval [0,1], the environment is said to be *Q-model.* On the other hand, if the inputs are arbitrary number in the closed line segment [0,1], the environment is known as *S-model,*
- Q is a finite set of states, $Q = \{q_1, \dots, q_s\}$,
- U is a finite set of outputs, $U = \{u_1, \dots, u_m\}$,
- F is the next state function

$$q(n+1) = F[y(n), q(n)], \tag{4}$$

- and G is the output function

$$u(n) = G[q(n)]. \tag{5}$$

– In general, the function F is stochastic and the function G may be deterministic of stochastic.

– Because of the stochastic nature in state transitions, stochastic automata are considered suitable for modeling learning systems.

– If the output of the automaton is $u_j, j = 1, 2, \dots, m$, the random environment generates a penalty with probability π_j or a nonpenalty with probability $(1 - \pi_j)$.

– The overall average penalty

$$I(n) = \sum_{i=1}^{m} P_i(n)\pi_i \tag{6}$$

Definition 2
If

$$\lim_{n\to\infty} \mathbb{E}\ \{I(n)\} < (1/m) \sum_{j=1}^{m} \pi_j, \tag{7}$$

the performance of the automaton is called expedient.

– Expediency being defined as the closeness of I to $I_{\min} = \min(\pi_1, \dots, \pi_m)$.

– Assume that $\pi_\beta = \min_i\{\pi_i\}$. Then, the optimal action of the stochastic automaton is u_β.

Definition 3
A reinforcement scheme is said to be optimal if

$$\lim_{n\to\infty} \mathbb{E}\left\{P_\beta(n)\right\} = 1 \tag{8}$$

Definition 4
A reinforcement scheme is said to be ϵ-optimal if

$$\lim_{\alpha\to 0} \lim_{n\to\infty} \mathbb{E}\left\{P_\beta(n)\right\} = 1 \tag{9}$$

where α is the learning rate.

- The above definition implies that ϵ-optimality ensures the learning property of stochastic automaton which is very close to optimality. From Definition 4 the following property can be derived.

  For an arbitrary positive number ϵ, there exists some parameter α_0 which ensures

$$\lim_{n\to\infty} \mathbb{E}\left\{P_\beta(n)\right\} \geq 1-\epsilon, \quad \text{for } |\alpha| < \alpha_0 \tag{10}$$

Definition 5
A reinforcement scheme is said to be absolutely expedient if

$$\mathbb{E}\left\{I(n+1)|P(n)\right\} < I(n) \tag{11}$$

- The definition of optimality and ϵ-optimality can be transformed to the definitions described by $I(n)$.

Definition 6
A reinforcement scheme is said to be optimal if

$$\lim_{n\to\infty} \mathbb{E}\left\{I(n)\right\} = \pi_\beta \tag{12}$$

Definition 7
A reinforcement scheme is said to be ϵ-optimal if

$$\lim_{\alpha\to 0} \lim_{n\to\infty} \mathbb{E}\left\{I(n)\right\} = \pi_\beta \tag{13}$$

where α is the learning rate.