SWITCHING MODELS FOR NON-STATIONARY RANDOM ENVIRONMENTS*

B. John Oommen+ and Hassan Masum
School of Computer Science
Carleton University
Ottawa, Canada : K1S 5B6.

ABSTRACT

Learning automata are stochastic finite state machines that attempt to learn the characteristic of an unknown random environment with which they interact. The fundamental problem is that of learning, through feedback, the action which has the highest probability of being rewarded by the environment. The problem of designing automata for stationary environments has been extensively studied. When the environment is non-stationary, the question of modelling the non-stationarity is, in itself, a very interesting problem. In this paper, we generalize the model used in [14,15] to present three models of non-stationarity. In the first two cases, the non-stationarity is modelled by a homogeneous Markov chain governing the way in which the characteristics change. The final model considers the more general case when the transition matrix of *this* chain itself changes with time in a geometric manner. In each case we have analyzed the stochastic properties of the resultant switching environment. The question of analyzing the various learning machines when interacting with these environments introduces an entire new avenue of open research problems. We are currently investigating how the three models introduced here (and in particular, the time varying model) are applicable to modelling telephone traffic.

* Partially supported by the Natural Sciences and Engineering Research Council of Canada. A preliminary version of some of these results can be found in the Proceedings of the 1993 IEEE International Conference on Systems, Man and Cybernetics, October 1993, Le Touquet, FRANCE.

+ Senior Member, IEEE.

I. INTRODUCTION

A Learning Automaton (LA) is a finite state machine which interacts with a random environment attempting to determine the best action which the latter offers. Such automata have also been used in the literature to model biological learning systems. The learning is achieved by interacting with the environment and processing its responses to the actions that are chosen. The concept of the LA was initially presented by Tsetlin [14,15] in his pioneering work in which he attempted to model biological learning. These automata find their applications in various fields such as game playing, parameter optimization, statistical decision making, telephone routing, adaptive queueing and object partitioning [1,3,5,6-8,10-12]. Since the literature on these automata is extensive, we refer the reader to two excellent books on the field by Lakshmivarahan [3] and Narendra and Thathachar [6] for a review of the families and applications of learning automata.

The learning process of an automaton can be described as follows: The automaton is offered a set of actions by the environment with which it interacts, and it is constrained to choose one of these actions. When an action is chosen, the automaton is either rewarded or penalized by the environment with a certain probability. A learning automaton is one which learns the optimal action, which is the action which has the minimum penalty probability. Hopefully, the automaton will eventually choose this action more frequently than other actions.

Variable Structure Stochastic Automata (VSSA) were developed by Varshavskii and Vorontsova. For these automata, the learning process is generalized so that the state transition probabilities and the action selecting probabilities evolve with time [3,6]. The automaton is simplified in the sense that each state now corresponds uniquely to a particular action. Hence while in state ϕ_i the automaton always picks one action α_i from a finite set A ; consequently, the set of states is redundant. Thus, what remains is the set of actions (or output from the automaton), the set of inputs B and a learning algorithm T. The element '0' $\in B$, is typically regarded as a reward, and the response '1' $\in B$ as a penalty -- such an environment is called a P-environment in the literature pertaining to learning systems. The learning algorithm operates on a probability vector, $\mathbf{P}(n) = \left(p_1(n), p_2(n) , . . .p_r(n)\right)$ called the action probability vector, where $p_i(t) = \mathbf{Pr}\{\alpha(t) = \alpha_i\}$ is the probability that the automaton will select action α_i at time n. In the case when the probability space is discretized, various discretized VSSA with superior properties [9] have been devised; when the automata also utilize information from running estimates of the penalty probabilities, continuous and discretized estimator algorithms have been reported [4,12].

VSSA are analyzed from the point of view that their probability of choosing an action at a given time represents a discrete Markovian process. The probability that an action may be rewarded can remain stationary or non-stationary depending on the environment; VSSA have been developed that are suitable for each situation. Many varieties of VSSA which possess absorbing barriers have been documented [3,6]; ergodic schemes have also been studied. The latter VSSA can go into and out of any state an unlimited number of times, and they possess the property that their limiting behaviour is independent of their initial state. In non-stationary environments, since the optimal action may change with time, an ergodic VSSA which can follow it is a suitable choice. In contrast, in stationary environments, automata with absorbing barriers are preferred because they can be made to converge to the optimal action with a probability as close to unity as desired.

A stochastic environment is defined as a 3-tuple : <*A, B,* **C**>. The sets *A* and *B* are the set of actions and the set of responses respectively. The set **C** = [$c_1, c_2, \ldots, c_r$] is the vector of penalty probabilities, where $c_i(t) = \mathbf{Pr}\{\ \beta(t) = 1 \mid \alpha(t) = \alpha_i\ \}$.

Most of the literature on adaptive learning deals with the problem of designing LA which interact with stationary environments. However, in many real life situations, (for example, in a telephone network [6-8,13]) it is quite unreasonable for us to assume that the probabilities of being rewarded are stationary. This is the primary focus of this paper.

In this paper we shall consider the problem of modelling the non-stationarity of the random environment itself. The types of models for such environments are few. Tsetlin [14,15], who pioneered the work in this area, presented the first model for non-stationarity. In his work, Tsetlin devised a model in which the non-stationary environment was composed of an ensemble of stationary environments -- in this case the non-stationarity itself was assumed to be a consequence of a switching process. Thus the non-stationary environment $\mathfrak{E}$ is fully defined by a set of stationary environments $\{E_1, \ldots, E_H\}$ and a switching matrix, T, whose arbitrary element $T_{i,j}$ represents the probability that the LA interacts with E_j at the next time instant if it was interacting with E_i at the present time instant.

Other results in the literature model the non-stationarity differently. In [7] the penalty probabilities are modelled so as to vary with time and to be simultaneously dependent on the action chosen by the LA. This model was later generalized for the case when the penalty probabilities were functions of the action probabilities themselves [13]. Other models for hierarchical automata and for the case when the optimal action is always the same have also been studied in the literature [6]. A complete survey of these results is found in the excellent treatise of Narendra and Thathachar [6] and is not repeated here in the interest of brevity. A variety of new models for non-stationary environments which invoke queries in data retrieval systems have been catalogued [16].

In this paper we shall consider the model for the environment as a generalization of the model presented by Tsetlin [14,15]. The aim is to study various scenarios that are encountered when the environment switching matrix is constrained. We shall consider two cases when the matrix T itself is stationary. In the first scenario, we assume that the probability, $T_{i,j}$, of switching from E_i to E_j is dependent on the index 'i' *from* which the switching occurs. Such a model will be called a "Source Oriented" Model. In the second case, we shall consider the scenario when the probability

$T_{i,j}$ is dependent on the index 'j' *to* which the switching occurs. This model will be called the "Destination Oriented" Model. In both these cases, we will derive the asymptotic probabilities of the automaton interacting with the respective environments. The final model which we will present will be the general one in which the non-stationarity itself is time varying, but "moving" toward a fixed point. The time-varying nature of this non-stationarity is achieved by modelling T to *itself* vary with time geometrically. In this case we shall derive an expression for $q_i(n)$, the probability of the automaton interacting with the environment E_i at any time instant 'n'.

The question of analyzing the various families of learning machines (fixed structure, variable structure, and discretized) when interacting with these environments leads to a host of extremely interesting open research problems. Besides this, we believe that the three models introduced here (and in particular, the time varying model) lend themselves to modelling telephone traffic. We are currently investigating this.

Our formulation is also applicable for a wide class of time varying Markov chains. Thus, apart from the learning applications of the various models which we have presented, we believe that the models also have potential applications in various other fields such as in econometric modelling. Consider the case when a purchaser has opted to buy the stock in a company X_i chosen from a set of companies $\mathfrak{X}$. At the next time instant, he may choose to sell his stock X_i and buy stock in X_j. If the probability of him moving from X_i to X_j is dependent on i, the analysis which we have presented for the "Source Oriented" Model of transition will be applicable to evaluating the asymptotic probability of him choosing the stock of the various companies in $\mathfrak{X}$. Similarly, if the probability of him moving from X_i to X_j is dependent on j, the analysis which we present for the "Destination Oriented" Model of transition will be applicable to calculating the corresponding asymptotic probabilities. Finally, if the probability of switching from one company to another itself changes with time "in a slow manner", our analysis of the time varying case will be relevant.

II. TIME INVARIANT SWITCHING ENVIRONMENTS

Throughout this paper we assume that we are dealing with $\mathfrak{E} = \{E_1,\ldots,E_H\}$, a set of H environments. The non-stationarity of the environment is modelled so as to have been obtained by a switching arrangement which is itself Markovian and governed by an ergodic Markov matrix T. Thus, the system switches from E_i to E_j with a probability $T_{i,j}$. The case that has been studied extensively in the literature [14,15] is the case when :

$$\begin{array}{lcll} T_{i,j} & = & \delta & \text{if } i \neq j \\ & = & 1 - (H-1)\delta & \text{if } i=j. \end{array}$$

where $\delta < 1/H$. Notice that in this case δ represents the average frequency of switching, and that the switching arrangement switches from E_i to E_j with a probability which is independent of both the indices 'i' and 'j'. Notice too that a small value of δ implies a slowly varying environment. It is easily verified that the stationary (asymptotic) probabilities of the Markov chain are [1/H,...,1/H] for all permissible values of δ, and that the average length of the interval over which the environment remains in any E_i is determined by δ (in fact, it will just be $1/\delta$). Thus, although the mean time for which the LA interacts with any particular environment (prior to switching) is $1/\delta$, the LA interacts with each environment asymptotically with a probability 1/H which is independent of δ. The performance of some fixed structure LA operating in such an environment has been analyzed in [14,15].

We shall now consider the Source and Destination Oriented Models of switching.

2.1 The Source Oriented Model:

The first model which we introduce is the one in which the probability of the system switching from E_i to E_j at any time instant depends on the environment from which it is migrating. We shall refer to this model as the Source Oriented Model. Thus, this probability $T_{i,j}$ will be :

$$\begin{aligned} T_{i,j} &= s_i && \text{if } i \neq j \\ &= 1 - (H-1)s_i && \text{if } i = j. \end{aligned}$$

Note that if each s_i equals δ the model reduces to the switching environment model studied in [14,15]. Also, note that the sum of the s_i's does not have to be unity and that the rate of switching of the environment increases with s_i.

We shall now derive the asymptotic probability of the system being in any environment E_i.

Theorem I.

For the Source Oriented Model of environment switching, the asymptotic probability[1] of the LA to be interacting with environment E_i is given by q_i^*, where

$$q_i^* = \frac{1/s_i}{\sum_{j=1}^{H} (1/s_j)} \tag{1}$$

Proof :

Consider the Source Oriented Model for environment switching. If at time 'n' the LA was interacting with environment E_i then, by definition, it would interact with environment E_j at 'n+1' with probability s_i. Thus the mechanism which determines the next-environment function obeys the H x H Markov matrix T given below :

[1] Note that this probability is, strictly speaking, the asymptotic probability of the Environment being in *its* i^{th} state E_i, and has nothing to do with the automaton or the type of automaton it is interacting with.

$$T = \begin{bmatrix} 1\text{-}(H\text{-}1)s_1 & s_1 & . \; . \; . & s_1 \\ s_2 & 1\text{-}(H\text{-}1)s_2 & . \; . \; . & s_2 \\ . & . & . \; . \; . \; . & . \\ s_H & s_H & . \; . \; . & 1\text{-}(H\text{-}1)s_H \end{bmatrix}$$

with $0 < s_j < 1/(H\text{-}1)$.

Clearly T represents a single closed communicating class whose periodicity is unity. Hence the chain is ergodic, and the limiting probability vector is given by the eigenvector of T^T corresponding to the eigenvalue unity. Let this vector be $Q^* = [q_1^*, ..., q_H^*]$. Then, Q^* satisfies :

$$Q^* = T^T Q^*. \qquad (2)$$

Expanding the k^{th} row of this equation yields :

$$q_k^* = \sum_{j \neq k} s_j \, q_j^* + [1\text{-}(H\text{-}1)s_k] \, q_k^*$$

This leads us to the equality that

$$H \, s_k \, q_k^* = \sum_{j=1}^{H} s_j \, q_j^*$$

Since the right hand side of the above is independent of k, it can be subsumed in the normalizing constant. It is thus easily verified that q_i^* is inversely proportional to s_i^*. Consequently, the values of q_i^* that satisfy (1) also obey (2). Hence the theorem. ●●●

If c_i^j is the penalty probability associated with action α_i when the environment in question is E_j, then the asymptotic average penalty that the LA would receive if it were choosing actions randomly is M_o, where,

$$M_o = \sum_{i=1}^{r} \sum_{j=1}^{H} q_j^* \, c_i^j$$

Thus, if $E[M(n)]$ is the average penalty obtained by the LA at time 'n', the LA will be deemed to be expedient [6] if :

$$\lim_{n \to \infty} E[M(n)] < M_o.$$

2.2 The Destination Oriented Model:

Analogous to the Source Oriented Model, the probability of the system switching from E_i to E_j at any time instant could depend on the environment to which it is migrating. Such an environment model will be referred to as the Destination Oriented Model. The probability of the system switching from E_i to E_j will now be specified by a quantity s_j whenever $i \neq j$. More explicitly, the switching matrix will be specified by a matrix T, where if

$$S = \sum_{j=1}^{H} s_j \qquad (3)$$

the probability $T_{i,j}$ will be :

$$\begin{aligned} T_{i,j} &= s_j && \text{if } i \neq j \\ &= 1-(S-s_j) && \text{if } i = j. \end{aligned}$$

As in the Source Oriented Model, note that if each s_j equals δ the model reduces to the switching environment model studied in [14,15] ; also, a larger value for s_j implies a faster switching system.

We shall now derive the asymptotic probability of the system being in any environment E_i.

Theorem II.

For the Destination Oriented Model of environment switching, the asymptotic probability[2] of the LA to be interacting with environment E_i is given by q_i^*, where

$$q_i^* = \frac{s_i}{\sum_{j=1}^{H} s_j} \qquad (4)$$

Proof :

Consider the Destination Oriented model defined above. The switching mechanism which determines the next-environment function obeys the H x H Markov matrix T given below :

$$T = \begin{bmatrix} 1-(S-s_1) & s_2 & . . . & s_H \\ s_1 & 1-(S-s_2) & . . . & s_H \\ . & . & . . . . & . \\ s_1 & s_2 & . . . & 1-(S-s_H) \end{bmatrix}$$

In order to ensure that T is a stochastic matrix, we will require that for all i,

$$0 < s_i < 1 , \qquad \text{and}$$

$$0 < S - \min\{s_i\} < 1.$$

As before T represents an ergodic Markov chain. The asymptotic probability vector is given by the eigenvector of T^T corresponding to the eigenvalue unity. Let this vector be $Q^* = [q_1^*, ..., q_H^*]$. Then, Q^* satisfies :

$$Q^* = T^T Q^*. \qquad (5)$$

Expanding the kth row of this equation yields :

$$q_k^* = \sum_{j \neq k} s_k \, q_j^* + [1-(S-s_k)] \, q_k^*$$

This leads us to the equality that

[2] As in Theorem I, note that this probability is the asymptotic probability of the Environment being in *its* i^{th} state E_i, and has nothing to do with the automaton or the type of automaton it is interacting with.

$$q_k^* = s_k \sum_{j=1}^{H} q_j^* + q_k^* - S\, q_k^*.$$

Since q_k^* cancels from both sides of the equation and since Q^* is itself a probability vector it is easily verified that q_i^* is directly proportional to s_i^*.

Hence the values of q_i^* that satisfy (4) also obey (5), and the theorem is proved. ●●●

In this case, if c_i^j is the penalty probability associated with action α_i when the environment in question is E_j, the asymptotic average penalty that the LA would receive if it were choosing actions randomly is M_o, where,

$$M_o = \sum_{i=1}^{r} \sum_{j=1}^{H} q_j^* \, c_i^j$$

where q_j^* would be defined by (4). Again, the LA will be deemed to be expedient [6] if :

$$\lim_{n\to\infty} E[M(n)] < M_o.$$

We shall now consider the more important and far more interesting case occurring when the switching matrix itself is time varying.

III. TIME VARYING SWITCHING ENVIRONMENTS

It is easy to see that both the Source Oriented and the Destination Oriented Models are generalizations of the special case analyzed in detail by Tsetlin [14,15] except that, in our study, additional constraints have been imposed on the switching process. In this section, we shall consider the more general case in which the switching matrix itself is time varying. Thus, at time 'n', the probability of the system switching from E_i to E_j will be specified by $T_{i,j}(n)$. The general analysis of this system seems to be intractable, and must be handled on a case-by-case basis depending on the fashion in which T varies with time. In this study we shall present the scenario in which $T_{i,j}(n)$ is a function of 'n', where this variation *itself* decreases with time[3]. Thus, as time proceeds indefinitely, we assume that the switching pattern stabilizes.

To model the switching pattern, we generalize the Destination Oriented Model by assuming that $T_{i,j}(n)$, the probability of switching from E_i to E_j, decreases geometrically. More precisely, the model assumes that the time varying switching matrix is specified as the transpose[4] of :

[3] If the time-varying nature of the switching pattern does not stabilize, the question of the *existence* of a final solution itself remains open. The fact that such variations in environments stabilize is well known in telephony. Traffic patterns change drastically on days such as Christmas etc., but once they have changed they tend to stabilize.

[4] Note that in this case, for ease of notation used later in this section, we have defined T(n) as the *transpose* of the Markov transition matrix instead of as the Markov transition matrix itself.

$$T(n) = \begin{bmatrix} 1-a^n(S-s_1) & a^n s_1 & . \; . \; . & a^n s_1 \\ a^n s_2 & 1-a^n(S-s_2) & . \; . \; . & a^n s_2 \\ . & . & . \; . \; . \; . & . \\ a^n s_H & a^n s_H & . \; . \; . & 1-a^n(S-s_H) \end{bmatrix}, \qquad (6)$$

where $0 < a < 1$, and as in Section 2.2,

$$S = \sum_{j=1}^{H} s_j,$$

with $0 < s_j < 1$, and $0 < S - \min\{s_j\} < 1$.

Note that the variable keeping track of time is 'n', and that the parameter characterizing the variation with time is 'a'. As n increases, notice that the term a^n in each entry of the above matrix will approach zero since a is less than unity. Thus T(n) will approach an identity matrix as n increases indefinitely. This implies that the switching will eventually cease.

We now derive the properties of this system by first proving the following lemmas.

Lemma I.

Let A be any H x H matrix with H linearly independent eigenvectors. Let K be the matrix which diagonalizes[5] A. Let $B(n) = \mathbf{I} + x_nA$, where $\mathbf{I}$ is the Identity matrix and $\{x_n\}$ is a sequence of constants. Then B(n) can be diagonalized by the same matrix K for all n.

Proof :

Since K diagonalizes A, $K^{-1}AK$ is some diagonal matrix $\mathbf{D}$. But:

$$\begin{aligned} K^{-1} B(n) K &= K^{-1} (\mathbf{I} + x_nA) K \\ &= K^{-1} \mathbf{I} K + K^{-1} (x_nA) K \\ &= \mathbf{I} + x_nK^{-1} A K \quad = \quad \mathbf{I} + x_n\mathbf{D} \end{aligned}$$

which is also a diagonal matrix. Hence the lemma. ●●●

We shall use the above lemma to analyze the properties of T(n). The eigenvector properties of T(n) for the case when H is 2 have been derived in [11]. In the latter case, since the authors were dealing with a 2-dimensional probability vector, the computations were greatly simplified. In the present case, since T(n) is an H x H matrix, the corresponding proofs for its diagonalization are far more involved. We shall first find the matrix K which diagonalizes T(n) for all n.

[5] Since we shall be invoking the diagonalizing properties alluded to in Lemma I, we would like to remind the reader that a necessary and sufficient condition for any H x H matrix to be diagonalizable is that it should possess H linearly independent eigenvectors.

Lemma II.

The matrix, K, which has the form :

$$K = \begin{bmatrix} s_1/s_H & -1 & -1 & . & . & -1 \\ s_2/s_H & 0 & 0 & . & . & 1 \\ . & . & . & . & . & . \\ s_{H-1}/s_H & 0 & 1 & . & . & 0 \\ 1 & 1 & 0 & . & . & 0 \end{bmatrix}. \qquad (7)$$

diagonalizes T(n) for all values of n.

Proof :

Let us decompose T(n) to be of the form :

$$T(n) = \mathbf{I} + x_nA.$$

Then the matrix A is the time invariant matrix :

$$A = \begin{bmatrix} -(S-s_1) & s_1 & . \; . \; . & s_1 \\ s_2 & -(S-s_2) & . \; . \; . & s_2 \\ . & . & . \; . \; . \; . & . \\ s_H & s_H & . \; . \; . & -(S-s_H) \end{bmatrix}$$

where $S = \sum_{j=1}^{H} s_j$ and $0 < s_j < 1$.

To find the diagonalizing matrix for A, we need to find its H eigenvectors. It is easily verified that the following vectors are indeed eigenvectors for the matrix A :

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \cdots \\ \cdots \\ s_{H-1} \\ s_H \end{bmatrix}, \text{ and } \begin{bmatrix} -1 \\ 0 \\ 0 \\ \cdots \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ \cdots \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ \cdots \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} -1 \\ 1 \\ 0 \\ \cdots \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

These eigenvectors are clearly linearly independent, and thus the matrix consisting of the eigenvectors as columns diagonalizes A. Normalizing the first column by dividing by s_H yields the matrix K as :

$$K = \begin{bmatrix} s_1/s_H & -1 & -1 & . & . & -1 \\ s_2/s_H & 0 & 0 & . & . & 1 \\ . & . & . & . & . & . \\ s_{H-1}/s_H & 0 & 1 & . & . & 0 \\ 1 & 1 & 0 & . & . & 0 \end{bmatrix}.$$

By Lemma I, this same matrix K which diagonalizes A will also diagonalize T(n) for all n, since the multiplying factor x_n is nothing but a^n. Hence the result. ●●●

To further analyze the diagonalization process, we shall now evaluate the diagonalized form of T(n). It is well known that the diagonalized matrix will be a diagonal matrix with the eigenvalues on the main diagonal. Thus, to evaluate the diagonalized form of any particular T(n), we need to find its eigenvalues. Interestingly enough, although the same matrix of eigenvectors diagonalizes T(n) for all 'n', the corresponding eigenvalues will differ. The following lemma specifies the eigenvalues.

Lemma III.

The eigenvalues of T(n) are unity and (1- a^nS). Whereas the former eigenvalue has multiplicity unity, the latter has multiplicity H-1.

Proof :

It is easily verified that the columns of K are eigenvectors of T(n). Consider the first eigenvector $\mathbf{v}_1$. For $\mathbf{v}_1$ we have :

$$T(n)\,\mathbf{v}_1 \quad = \begin{bmatrix} 1-a^n(S-s_1) & a^n s_1 & . \; . \; . & a^n s_1 \\ a^n s_2 & 1-a^n(S-s_2) & . \; . \; . & a^n s_2 \\ . & . & . \; . \; . & . \\ a^n s_H & a^n s_H & . \; . \; . & 1-a^n(S-s_H) \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \cdots \\ \cdots \\ s_{H-1} \\ s_H \end{bmatrix}$$

$$= \begin{bmatrix} (1-a^n(S-s_1))s_1 + a^n s_1 s_2 + \ldots + a^n s_1 s_H \\ a^n s_2 s_1 + (1-a^n(S-s_2))s_2 + \ldots + a^n s_2 s_H \\ \cdots \\ a^n s_H s_1 + a^n s_H s_2 + \ldots + (1-a^n(S-s_H))s_H \end{bmatrix}$$

$$= \begin{bmatrix} (1-a^n(S-s_1))s_1 + a^n s_1(S-s_1) \\ (1-a^n(S-s_2))s_2 + a^n s_2(S-s_2) \\ \cdots \\ (1-a^n(S-s_H))s_H + a^n s_H(S-s_H) \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \cdots \\ \cdots \\ s_{H-1} \\ s_H \end{bmatrix} = \mathbf{v}_1$$

which demonstrates that unity is an eigenvalue.

Consider now the k^{th} eigenvector $\mathbf{v}_k$ ($2 \le k \le H$). For it we have:

$$T(n)\,\mathbf{v}_k \quad = \begin{bmatrix} 1-a^n(S-s_1) & a^n s_1 & . \; . \; . & a^n s_1 \\ a^n s_2 & 1-a^n(S-s_2) & . \; . \; . & a^n s_2 \\ . & . & . \; . \; . & . \\ a^n s_H & a^n s_H & . \; . \; . & 1-a^n(S-s_H) \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -(1-a^n(S-s_1)) + a^n s_1 \\ -a^n s_2 + a^n s_2 \\ \cdots \\ a^n s_k + (1-a^n(S-s_k)) \\ \cdots \\ -a^n s_H + a^n s_H \end{bmatrix} = \begin{bmatrix} -(1-a^n S) \\ 0 \\ \cdots \\ (1-a^n S) \\ \cdots \\ 0 \end{bmatrix}$$

$$= (1-a^n S)\,\mathbf{v}_k$$

This demonstrates that 1- a^n S is an eigenvalue of T(n). Since this is true for all ($2 \le k \le H$) it is clear that this eigenvalue has multiplicity H-1. Hence the lemma. ●●●

We now come to the final result which specifies the exact closed form expression for the probability of the system being in environment E_i at any time instant 'n'. This is done by utilizing the above diagonalization results. In [11], the case for the 2 x 2 scenario is solved explicitly. The present analysis generalizes these results for the H x H case.

Theorem III.

Let the initial probabilities of being in any environment be given by the vector of probabilities:

$$Q(0) = [q_1(0),...,q_H(0)]^T,$$

where $\sum_{j=1}^{H} q_j(0) = 1$. Then, for all time instants 'n', if $\xi_n = \prod_{j=0}^{n-1} (1-a^j S)$, the probability vector of the system being in the various environments is given by Q(n), where Q(n) is :

$$Q(n) = (1/S) * \begin{bmatrix} s_1 + \xi_n \left(q_1(0)\, S - s_1\right) \\ s_2 + \xi_n \left(q_2(0)\, S - s_2\right) \\ \cdots \\ s_H + \xi_n \left(q_H(0)\, S - s_H\right) \end{bmatrix}.$$

Proof :

Let the diagonalized form of T(n) be the diagonal matrix D(n). Then, by the above lemmas, D(n) has the form $D(n) = K^{-1}\, T(n)\, K$. Indeed, D(n) itself is the diagonal matrix with 1 in the upper left position and 1 - a^nS elsewhere along the main diagonal. Thus :

$$D(n) = \begin{bmatrix} 1 & 0 & . \; . \; . & 0 \\ 0 & 1-a^n S & . \; . \; . & 0 \\ . & . & . \; . \; . & . \\ 0 & 0 & . \; . \; . & 1-a^n S \end{bmatrix}$$

The Markov property of the chain forces Q(n) to satisfy :

$$Q(k+1) = T(k)\, Q(k). \qquad (8)$$

By invoking (8) repeatedly we have :

$$Q(n) = T(n-1).T(n-2) \ldots T(0)\, Q(0).$$

Using now the diagonal form of T(j) we get :

$$
\begin{aligned}
Q(n) &= \prod_{j=0}^{n-1} K\ D(j)\ K^{-1}\ Q(0) \\
&= K\ (\prod_{j=0}^{n-1} D(j))\ K^{-1}\ Q(0) \\
&= K\ \Xi_n\ K^{-1}\ Q(0) \qquad (9)
\end{aligned}
$$

where $\xi_n = \prod_{j=0}^{n-1} (1-a^j S)$, and Ξ_n is the H x H diagonal matrix :

$$
\Xi_n \quad = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & \xi_n & . & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \xi_n \end{bmatrix}. \qquad (10)
$$

Since the general form of the matrix K is as given by Lemma II, we are now left with the task of computing K^{-1}. By systematically computing the inverse (after much tedious algebra) it can be seen and easily verified that the general form of K^{-1} is :

$$
K^{-1} = (1/S) * \begin{bmatrix} s_H & s_H & . & . & s_H & s_H \\ -s_H & -s_H & . & . & -s_H & S-s_H \\ -s_{H-1} & -s_{H-1} & . & . & S-s_{H-1} & -s_H \\ . & . & . & . & . & . \\ -s_2 & S-s_2 & . & . & -s_2 & -s_2 \end{bmatrix}.
$$

We evaluate (9) by multiplying the matrices in a straightforward fashion to get

$$
K\,\Xi_n\,K^{-1} = (1/S) * \begin{bmatrix} s_1+s_2\xi_n +\ldots+s_H\xi_n & s_1-s_1\xi_n & . & . & . & s_1-s_1\xi_n \\ s_2-s_2\xi_n & s_1\xi_n+s_2 +\ldots+s_H\xi_n & . & . & . & s_2-s_2\xi_n \\ . & . & . & . & . & . \\ s_H-s_H\xi_n & s_H-s_H\xi_n & . & . & . & s_1\xi_n+s_2\xi_n +..+s_H \end{bmatrix}
$$

which simplifies to yield

$$
K\,\Xi_n\,K^{-1} = (1/S) * \begin{bmatrix} s_1+\xi_n(S-s_1) & s_1(1-\xi_n) & . & . & . & s_1(1-\xi_n) \\ s_2(1-\xi_n) & s_2+\xi_n(S-s_2) & . & . & . & s_2(1-\xi_n) \\ . & . & . & . & . & . \\ s_H(1-\xi_n) & s_H(1-\xi_n) & . & . & . & s_H+\xi_n(S-s_H) \end{bmatrix}.
$$

Using the form of Q(0) as specified by the statement of the theorem yields :

$$Q(n) = (1/S) * \begin{bmatrix} s_1 + \xi_n\big(q_1(0)\,(S\text{-}s_1) - s_1(1\text{-}q_1(0))\big) \\ s_2 + \xi_n\big(q_2(0)\,(S\text{-}s_2) - s_2(1\text{-}q_2(0))\big) \\ \cdots \\ s_H + \xi_n\big(q_H(0)\,(S\text{-}s_H) - s_H(1\text{-}q_H(0))\big) \end{bmatrix}$$

$$= (1/S) * \begin{bmatrix} s_1 + \xi_n\big(q_1(0)\,S - s_1\big) \\ s_2 + \xi_n\big(q_2(0)\,S - s_2\big) \\ \cdots \\ s_H + \xi_n\big(q_H(0)\,S - s_H\big) \end{bmatrix}.$$

and the theorem is proved. •••

We shall now give some concluding remarks regarding the above analysis. First of all it should be noted that the chain is time varying and so the traditional theorems for analyzing ergodic and absorbing chains cannot be utilized. Furthermore, on converging, the chain is "no more time varying", and is thus rendered absorbing since $T(\infty)$ becomes an identity matrix. Hence, the situation which we have here is an interesting "hybrid" whose non-steady-state solution can be computed easily by virtue of the diagonalizability of T(n). Note that in this case the long-term behaviour of the switching environments depends upon the value for ξ_n. In particular, the above formula for Q(n) shows that the asymptotic distribution of the probability mass will be wholly dependent upon the limiting value of ξ_n. If ξ_n tends to unity, then the vector $Q(\infty)$ will be very close to the vector Q(0). As opposed to this, if ξ_n tends to zero, the vector $Q(\infty)$ will approach a distribution which will be proportional to the corresponding values of the $\{s_j\}$.

In conclusion, we would like to remark that the sequence of coefficients that we have used in the definition of T(n), and subsequently for x_n in Lemma I is $\{1, a, a^2, a^3, ...\}$. Clearly, since we have not utilized the particular properties of this sequence except in the expression for ξ_n, the sequence can be generalized. Indeed, our results are applicable for any sequence $\{x_n\}$ that yields a convergent product for ξ_n.

IV. CONCLUSIONS

In this paper we have considered the paradigm of learning traditionally considered when stochastic automata are utilized to learn from a random environment. The problem of designing automata for stationary environments has been extensively studied in the literature. We have considered the problem of modelling non-stationary environments. The original switching model due to Tsetlin [14,15] has been specialized for two scenario and generalized in a third to present three new models of non-stationarity. In the first two cases, the non-stationarity is modelled by a homogeneous Markov chain which governs the way in which the environments change; the probabilities of transition either depend on the environment *from* which the transition takes place, or the environment *to* which the transition occurs. The final model considers the more general case when the environment switching transition matrix itself changes with time in a geometric manner. In each case we have analyzed the stochastic properties of the resultant switching environment.

The question of analyzing the various learning machines when interacting with these environments now introduces an entire new avenue of open research problems. We are also currently studying how the three models introduced here (and in particular, the time varying model) are applicable to modelling telephone traffic.

REFERENCES

[1] Baba, S., Soeda, S.T., and Sawaragi, Y., "An Application of Stochastic Automata to the Investment Game," *Int. J. Syst. Sci.*, vol. 11, no. 12, pp 1447-1457, Dec. 1980.

[2] Karlin, S., Taylor, H. M., *A First Course on Stochastic Processes*, Academic Press, 1974.

[3] Lakshmivarahan, S., *Learning Algorithms Theory and Applications*, Springer-Verlag, 1981.

[4] Lanctôt, J. K., *Discrete Estimator Algorithms: A Mathematical Model of Computer Learning*, M.Sc. Thesis, Department of Mathematics and Statistics, Carleton University, Ottawa, Canada, 1989.

[5] Meybodi, M.R., *Learning Automata and Its Application to Priority Assignment in a Queueing System With Unknown Characteristic*, Ph.D. Thesis, School of Electrical Engineering and Computing Sciences, University of Oklahoma, Norman, Oklahoma.

[6] Narendra, K.S., and Thathachar, M.A.L., *Learning Automata*, Prentice-Hall, 1989.

[7] Narendra, K.S., and Thathachar, M.A.L., "On the Behaviour of a Learning Automaton in a Changing Environment With Routing Applications", *IEEE Trans. on Syst. Man and Cybern.*, Vol. SMC-10, 1980, pp.262- 269.

[8] Narendra, K.S., Wright, E., and Mason, L.G., "Applications of Learning Automata to Telephone Traffic Routing", *IEEE Trans. on Syst. Man and Cybern.*, Vol. SMC-7, 1977, pp.785-792.

[9] Oommen, B.J., "Absorbing and Ergodic Discretized Two-Action Learning Automata", *IEEE Trans. on Syst. Man and Cybern.*, Vol. SMC-16, 1986, pp.282-296.

[10] Oommen, B. J., and Ma, D. C. Y., "Deterministic Learning Automata Solutions to the Equipartitioning Problem", *IEEE Transactions on Computers*, Vol. 37, January 1988, pp.2-14.

[11] Oommen, B.J., and Hansen E.R., "List Organizing Strategies using Stochastic Move-to-Front and Stochastic Move-to-Rear Operations", *SIAM Journal of Computing*, Vol. 16, No. 4, August 1987, pp. 705-716.

[12] Sastry, P.S., *Systems of Learning Automata: Estimator Algorithms Applications*, Ph.D. Thesis, Dept of Electrical Engineering, Indian Institute of Science, Bangalore, India, June 1985.

[13] Srikantakumar, P. R., and Narendra, K. S., "A Learning Model for Routing in Telephone Networks", *SIAM J. Control and Optimization*, Vol. 20, 1982, pp.34-57.

[14] Tsetlin, M.L., "On the Behaviour of Finite Automata in Random Media", *Automat. Telemekh.*(USSR), Vol.22, Oct. 1961, pp.1345-1354.

[15] Tsetlin, M.L., *Automaton Theory and the Modelling of Biological Systems*, New York and London, Academic Press, 1973.

[16] Valiveti, R.S. and Oommen, B.J., "The Move-to-Front List Organizing Heuristic for Non-Stationary Query Distributions", *Proceedings of the 1991 International Symposium on Computer and Information Sciences*, October/November 1991, Antalya, Turkey, pp. 105-114.