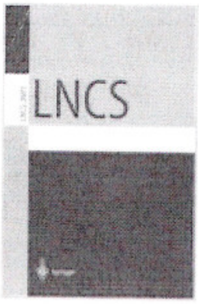
Chapter



Lecture Notes in Computer Science
Publisher: Springer-Verlag Heidelberg
ISSN: 0302-9743
Volume: Volume 2869 / 2003
Date: October 2003
Pages: 960 - 967
Title: Computer and Information Sciences - ISCIS 2003
ISBN: 3-540-20409-1

Soft Computing

A New Continuous Action-Set Learning Automaton for Function Optimization

Hamid Beigy[1] and M.R. Meybodi[1]

(1) Soft Computing Laboratory, Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran,

In this paper, we study an adaptive random search method based on learning automaton for solving stochastic optimization problems in which only the noise-corrupted value of objective function at any chosen point in the parameter space is available. We first introduce a new continuous action-set learning automaton (CALA) and theoretically study its convergence properties, which implies the convergence to the optimal action. Then we give an algorithm, which needs only one function evaluation in each stage, for optimizing an unknown function.

The references of this article are secured to subscribers.

A New Continuous Action Learning Automata Based Limited Fractional Guard Channel Policy

Hamid Beigy and M. R. Meybodi*

Soft Computing Laboratory
Computer Engineering and Information Technology Department
Amirkabir University of Technology
Tehran, Iran
{beigy, meybodi}@ce.aut.ac.ir

Abstract

In this paper, we first propose a new continuous action-set learning automaton and then give an adaptive and autonomous call admission algorithm, which uses the proposed learning automaton. The proposed algorithm minimizes blocking probability of new calls subject to constraint on dropping probability of handoff calls. The simulation results show that the performance of the proposed algorithm is close to the performance of limited fractional guard channel policy for which we need to know all traffic parameters in advance.

Keywords

Learning Automata, Continuous Action Learning Automata, Adaptive Call Admission Control, Guard Channel Policy, Limited Fractional Guard Channel Policy,

1. Introduction

Call admission policies control both blocking probability of new calls (B_n) and dropping probability of handoff calls (B_h) by putting some restrictions on the allocation of channels to the incoming calls. Since the dropping probability of handoff calls is more important than the blocking probability of new calls, call admission policies usually put restriction on acceptance of new calls. Assume that the given cell has C full duplex channels. Guard channel policy reserves a subset of channels allocated to the cell for sole use of handoff calls (say $C - T$ channels). These channels are called *guard channels*. In guard channel policy, when the channel occupancy exceeds certain threshold T, then the new calls are rejected until the channel occupancy goes below threshold T [1]. This policy accepts handoff calls as long as channels are available. As the number of guard channels increased, the dropping probability of handoff calls will be reduced while the blocking probability of new calls will be increased [2]. If the parameter B_h is considered, the guard channel policy gives very good performance, but the parameter B_n is degraded to great extent. In order to have more control on blocking probability of new calls and the dropping probability of handoff calls, *limited fractional guard channel policy* (LFG) is introduced [3]. The LFG policy uses an additional parameter π. The LFG policy is the same as the GC policy except when T channels are occupied in the cell, the LFG policy accepts new calls with probability π. It has been shown that there is an optimal threshold T^* and an optimal value of π^* for which the blocking probability of new calls is minimized subject to the hard constraint on the dropping probability of handoff calls [3]. Since the input traffic is not a stationary process and its parameters are unknown a priori, the optimal number of guard channels is different for different traffic. In such cases the *dynamic guard channel* policy in which the number of guard channels varies during the operation of the cellular network can be used.

The learning automaton (LA) is adaptive decision making device that operating in an unknown random environment and progressively improves its performance via a learning process. LAs are divided into two main groups: finite action-set LA (FALA) and continuous action-set LA (CALA) based on whether the action set is finite or continuous [4]. FALA has finite number of actions and has been studied extensively. For an $r-$action FALA, the action probability distribution is represented by an $r-$dimensional probability vector that updated by the learning algorithm. In many applications there is need to have large number of actions. The LA with too large number of actions

*This work is partially supported by Iranian Telecommunication Research Center (ITRC), Tehran, Iran.

converge too slowly. In such applications CALA, whose actions are chosen from real line, are very useful. For CALA, the action probability distribution is represented by a continuous function and this function is updated by learning algorithm at any stage. In the theory of LA, several algorithms for learning optimal parameters have been developed for many discrete and continuous parameters. Gullapalli proposed a generalized LA with continuous action set, which uses the context input as well as the reinforcement signal [5]. In [6], a continuous action-set LA is given and its convergence has been shown. This CALA is used for stochastic optimization and needs only two function evaluations in each iteration, irrespective of the dimension of the parameter space. In [7], a CALA, which is called *continuous action reinforcement learning automata* (CARLA) is given for adaptive control, but its behavior is unknown. Learning automata have been used successfully in many applications such as computer network [8], solving NP-Complete problems [9] and neural network engineering [10, 11] to mention a few.

In this paper, we first propose a new continuous action-set LA in which μ and σ are adjusted using reinforcement signal. This LA uses a Gaussian distribution $N(\mu, \sigma)$ for choosing its actions. Then we introduce an adaptive and autonomous call admission algorithm, which uses the proposed continuous action-set LA. This algorithm uses only the current channel occupancy of the given cell and dynamically adjusts the number of guard channels. The proposed algorithm minimizes the blocking probability of new calls subject to the constraint on the dropping probability of handoff calls. Since the LA starts its learning without any priori knowledge about its environment, the proposed algorithm does not need any a priori information about input traffic. One of the most important advantages of the proposed algorithm is that no status information will be exchanged between the neighboring cells. The exchange of such status information increase the performance of the proposed algorithm. The simulation results show that performance of the proposed algorithm is near to the performance of LFG policy that knows all the traffic parameters.

The rest of this paper is organized as follows. Section 2. presents the performance parameters of LFG policy. In section 3., a new continuous action-set LA is given and its behavior is studied. In section 4., an adaptive call admission control algorithm is given which uses the proposed continuous action-set LA. The simulation results is given in section 5. and section 6. concludes the paper.

2. Blocking Performance of LFG

In what follows, blocking performance of LFG policy is given. The blocking performance of LFG policy is computed based on assumptions: 1) The arrival process of new and handoff calls is poisson process with rate λ_n and λ_h, respectively. 2)The call holding time for both types of calls is exponentially distributed with mean μ^{-1}. 3) The time interval between two calls from a mobile host is much greater than mean call holding time. 4) Only mobile to fixed calls are considered. 5) The network is homogenous. The above first three assumptions have been found to be reasonable as long as the number of mobile hosts in a cell is much greater than the number of channels allocated to that cell. The fourth assumption makes our analysis easier and the fifth one lets us to examine the performance of a single network cell in isolation. Suppose that the given cell has a limited number of full duplex channels, C, in its channel pool. We define the state of a particular cell at time t to be the number of busy channels in that cell, which is represented by $c(t)$. The $\{c(t)|t \geq 0\}$ is a continuous-time Markov chain with states $0, 1, \ldots, C$. The state transition rate diagram of LFG policy is shown in figure 1.

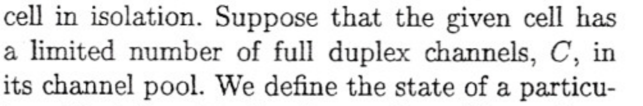


Fig. 1. Markov chain model of cell

Define the steady state probability

$$P_n = \lim_{t\to\infty} \text{Prob}[c(t) = n] \qquad n = 0, 1, \ldots, C. \tag{1}$$

Let $\lambda = \lambda_n + \lambda_h$, $\rho = \lambda/\mu$, and $\alpha = \lambda_h/\lambda$. In the steady state, the following expression can be derived for P_n.

$$P_n = \frac{\rho^n \prod_{i=1}^{n} \gamma_i}{n!} P_0 \tag{2}$$

where γ_i is the arrival rate in state i and P_0 is equal to

$$P_0 = \left[\sum_{k=0}^{C} \frac{\rho^k \prod_{i=1}^{k} \gamma_i}{k!}\right]^{-1} \tag{3}$$

Hence, the dropping probability of handoff calls for the LFG policy is equal to

$$B_h(C, T) = \frac{\rho^C \alpha^{C-T}}{C!}. \tag{4}$$

Similarly, the blocking probability of new calls is given by the following expression.

$$B_n(C,T) = \sum_{k=T+1}^{C} P_k + (1-\pi)P_T \qquad (5)$$

3. A New Continuous Action-set Learning Automaton

In this section, we introduce a new continuous action-set LA (CALA), which will be used later for determining the optimal parameter of LFG policy. For the proposed LA, we use the Gaussian distribution, $N(\mu, \sigma)$ for selection of actions, which is completely specified by the first and second order moments, μ and σ. Thus, unlike FALA, we only need to store the mean and variance parameters of the Gaussian distribution. The learning algorithm now updates the mean and variance of the Gaussian distribution at any instant using the evaluation signal β, obtained from the random environment. The evaluation signal, $\beta \in [0,1]$, is a noise-corrupted evaluation signal, which indicates a noise-corrupted observation of an unknown function $M(.)$ at the selected action. The evaluation signal β is a random variable whose distribution function coincides almost with the distribution $H(\beta|\alpha)$ that belongs to a family of distributions which depends on the parameter α. Let

$$M(\alpha) = \int_{-\infty}^{\infty} \beta(\alpha) dH(\beta|\alpha),$$

be a penalty function with bound $\mathcal{M}$ corresponding to this family of distributions. We assume that $M(.)$ is measurable and continuously differentiable almost everywhere. The CALA has to minimize $M(.)$ by observing $\beta(\alpha)$. Let (μ_n, σ_n) be the mean and the standard deviation of the Gaussian distribution that represents the action probability of the CALA at instant n. Through the learning algorithm, we ideally want that $\mu_n \to \mu^*$ and $\sigma_n \to 0$ as $n \to \infty$. The interaction between the CALA and the random environment takes place as iterations of the following operations. Iteration n begins by selection of an action α_n by the CALA. This action is generated as a random variable from the Gaussian distribution with parameters μ_n and σ_n. The selected action is applied to the random environment and the LA receives an evaluative signal $\beta(\alpha_n)$, which has the mean value $M(\alpha_n)$, from the environment. Then the LA updates the parameters μ_n and σ_n. Initially $M(.)$ is not known and it is desirable that with interaction of LA and the random environment, the μ and σ converges to their optimal values which results the minimum value of $M(.)$. The LA uses the following rule to update its parameters, thus generating a sequence of random variables μ_n and σ_n.

$$\begin{aligned} \mu_{n+1} &= \mu_n - a\beta(\alpha_n)\sigma_n(\alpha_n - \mu_n), \\ \sigma_{n+1} &= f(\sigma_n), \end{aligned} \qquad (6)$$

where a is learning rate and $f(.)$ is a function that produces a random sequence of σ_n (described later). The equation (6) can be written as

$$\mu_{n+1} = \mu_n - a\sigma_n^2 y_n(\alpha_n), \qquad (7)$$

where

$$y_n(\alpha_n) = \beta(\alpha_n)\left(\frac{\alpha_n - \mu_n}{\sigma_n}\right). \qquad (8)$$

An intuitive explanation for the above updating equations is as follows. We can view the fraction in equation (8) as the normalized noise added to the mean. Since a, β and σ all are positive, the updating equation changes the mean value in the opposite direction of the noise. If the noise is positive, then the LA should update its parameter, so that mean value increased and visa versa. Since $E[\beta|\alpha]$ is close to unity when α is far from its optimal value and it is close to zero when α is near to the optimal value, thus the LA updates μ with large steps when α is far from its optimal value and with small steps when α is close to its optimal value. This causes a finer quantization of μ near its optimal value and a grain quantization for points far away its optimal value. Thus, we can consider the learning algorithm as a random direction search algorithm with adaptive step sizes.

In what follows, we state the convergence of the proposed CALA in stationary environments. The convergence is proved based on the following assumption.

Assumption 1. The sequence of real numbers $\{\sigma_n\}$ is such that $\sigma_n \geq 0, \sum_{n=1}^{\infty} \sigma^3 = \infty$, and $\sum_{n=1}^{\infty} \sigma^4 < \infty$.

Assumption 2. There is a unique real number μ^* such that $M(\mu^*) = \min_\alpha(\alpha)$. Suppose that $\mathcal{M}$ be the bound on $M(\alpha)$. It is also assumed that $M(\alpha)$ has a finite number of minima inside a compact set and has bounded first and second derivatives with respect to α. Let $R(\alpha) = \frac{\partial M(\alpha)}{\partial \alpha}$ and $S(\alpha) = \frac{\partial M^2(\alpha)}{\partial \alpha^2}$ be the first and second derivative of $M(\alpha)$, respectively. Suppose that $\mathcal{R}$ and $\mathcal{S}$ be the bounds on R and S, respectively.

Assumption 3. Suppose that $R(\alpha)$ is linear near μ^*, that is $\sup_{\epsilon \leq |\alpha-\mu^*| \leq \frac{1}{\epsilon}} (\alpha - \mu^*)R(\alpha) > 0$, for all $\epsilon > 0$.

Assumption 4. Suppose that the noise in the reinforcement signal $\beta(.)$ has a bounded variance, that is

$$\mathrm{E}\left\{[\beta(\alpha) - M(\alpha)]^2\right\} \leq K_1\left[1 + (\alpha - \mu^*)^2\right] \quad (9)$$

for some real number $K_1 > 0$. Thus $\mathrm{E}\left\{\beta^2(\alpha)\right\}$ is bounded by a quadratic function of α for all α.

Given the above assumptions, the following theorem states the convergence of the proposed CALA.

Theorem 1. *Suppose that the assumptions 1-4 hold, μ_0 is finite and there is an optimal value of μ^* for μ. Then if μ_n and σ_n are evolved according to the given learning algorithm, then $\lim_{n\to\infty} \mu_n = \mu^*$ with probability 1.*

Proof. The proof of this theorem is given in [12].

4. An Adaptive Call Admission Control Algorithm

In this section, we introduce a new LA based algorithm to determine the optimal value of $T + \pi$ for LFG algorithm when parameters λ_n, λ_h, and μ are unknown and possibly time varying. Assume that the cell has C full duplex channels. Let $\alpha_n = T_n + \pi_n$ be the parameter of LFG algorithm at instant n and $\alpha^* = T^* + \pi^*$ be the optimal value of α, which minimizes the blocking probability of new calls subject to constraint that the dropping probability of handoff calls is at most p_h. Let α_n be in interval $[g_{\min}, g_{\max}]$, where $0 \leq g_{\min} \leq g_{\max} \leq C$. In the proposed algorithm, each base station uses a LA described in the previous section, to adjust α. The objective of our call admission control policy is to minimize the blocking probability of new calls $(B_n(C,\alpha))$ subject to the hard constraint on the dropping probability of handoff calls$(B_h(C,\alpha) \leq p_h)$. Since α_n must be in the interval $[g_{\min}, g_{\max}]$, the proposed LA cannot be applied directly and we use a projected version of it. In the projected version of the CALA, the constraint set H is $\{\mu | g_{\min} \leq \mu \leq g_{\max}\}$. Thus the updating equation for μ in the proposed LA is replaced by

$$\mu_{n+1} = \Pi_H\left(\mu_n - a\beta_n\sigma_n(\alpha_n - \mu_n)\right)),$$

where Π_H is the projection on to the constraint set H and is defined by the following relation.

$$\Pi_H(x) = \begin{cases} g_{\min} & \text{if } x < g_{\min} \\ x & \text{if } g_{\min} \leq x \leq g_{\max} \\ g_{\max} & \text{if } x > g_{\max} \end{cases}$$

The standard deviation of Gaussian distribution is updated according to the following equation.

$$\sigma_n = \frac{1}{\lfloor \frac{n}{2000} \rfloor^{1/3}}$$

Now, we describe the proposed adaptive call admission algorithm. Since the controlling handoff calls is not beneficial as it leads to idling of some channels, the handoff calls are accepted as long as channels are available and the proposed algorithm is used only when new calls arrive. When a new call arrives at the given cell, LA associated to that cell selects one of its actions, say $\alpha_n = T + \pi$. If the number of busy channels of the cell is less than T, then the incoming call is accepted; when the cell has T busy channels the call is accepted with a certain probability of π; otherwise the incoming call is blocked. The LA updates its action probability distribution on the arrival of the next new call. The objective of updating is to find an action α^*,which minimizes the blocking probability of new calls subject to the hard constraint on the dropping probability of handoff calls $(B_h(C\alpha) \leq p_h)$. In order to find the optimal pair (T^*, π^*), the following performance index is defined.

$$M(\alpha) = |B_h(C,\alpha) - p_h|. \quad (10)$$

From the above performance index, it is evident that there is a unique α^* for which $M(\alpha^*) = 0$ and in this point the blocking probability of new calls are also minimized. Thus the call admission control is defined as finding the zero of function $M(.)$. The proposed algorithm updates the parameters of the Gaussian distribution to find the optimal action. This updating is done in such a way that the dropping probability of handoff calls approximately equal to p_h. Since the blocking probability of new calls is a decreasing function of $T + \pi$ and performance index $M(.)$ has only one minimum, thus the proposed algorithm minimizes the blocking probability of new calls while maintaining the level of the dropping probability of handoff calls.

5. Simulation Results

In this section, we compare performance of LFG [3] and proposed algorithm. The results of simulations are summarized in table 1. The simulation is based on the single cell of homogenous cellular network system. In such network, each cell has 8 full duplex channels. In the simulations, new call arrival rate is 30 calls per minute, channel holding time is set to 6 seconds, and the handoff call traffic is varied between 2 calls per minute to 20 calls per minute. The results listed in table 1 are obtained by averaging 10 runs from 2,000,000 seconds simulation of each algorithm. The objective is to minimize the blocking probability of new calls subject to the constraint that the dropping probability of handoff calls is less than 0.01. The optimal parameters of LFG policy is obtained by algorithm given in [3].

Table 1. The simulation results.

		LFG		Proposed Algorithm	
Case	λ_h	B_n	B_h	B_n	B_h
1	2	0.031609	0.023283	0.050722	0.010582
2	4	0.051414	0.020675	0.110385	0.010228
3	6	0.071632	0.018707	0.085284	0.009839
4	8	0.092138	0.016706	0.106879	0.010895
5	10	0.114445	0.015572	0.142426	0.010198
6	12	0.147902	0.014044	0.188929	0.00936
7	14	0.204217	0.012675	0.224772	0.009875
8	16	0.250642	0.011554	0.261856	0.009911
9	18	0.294441	0.010877	0.312300	0.010312
10	20	0.384157	0.010182	0.392528	0.010660

By inspecting table 1, it is evident that 1) the constraint $B_h(C, \alpha) \leq p_h$ is maintained by the proposed algorithm while it is not for the LFG policy. 2) Performance of the proposed algorithm (which doesn't need any a priori information) is close to the performance of LFG policy for which we need to know all traffic parameters. One reason for such difference is due to the transient behavior of the proposed algorithm. Since, $B_n(C, \alpha)$ and $B_h(C, \alpha)$ in the early stages of simulation are far from their desired value, they affect the long-time calculation of performance parameters. However, such effect can be removed by excluding transient behaviors of proposed algorithm for the calculation of $B_n(C, \alpha)$ and $B_h(C, \alpha)$.

6. Conclusions

In this paper, a new continuous action-set learning automaton is introduced and its convergence is studied. We stated convergence theorem that implies form of optimal performance for CALA. Then a call admission control algorithm is given which uses the proposed LA. The simulation results show the power of the proposed algorithm. The proposed algorithm has the following advantages: 1) Its performance is very close to the performance of the LFG policy. 2) The upper bound of dropping probability maintained by the proposed algorithm is optimal when all information are available. 3) The proposed algorithm doesn't need any information about the input traffic.

References

[1] D. Hong and S. Rappaport, "Traffic Modelling and Performance Analysis for Cellular Mobile Radio Telephone Systems with Priotrized and Nonpriotorized Handoffs Procedure," *IEEE Transactions on Vehicular Technology*, vol. 35, pp. 77–92, Aug. 1986.

[2] S. Oh and D. Tcha, "Priotrized Channel Assignment in a Cellular Radio Network," *IEEE Transactions on Communications*, vol. 40, pp. 1259–1269, July 1992.

[3] R. Ramjee, D. Towsley, and R. Nagarajan, "On Optimal Call Admission Control in Cellular Networks," *Wireless Networks*, vol. 3, pp. 29–41, 1997.

[4] M. A. L. Thathachar and P. S. Sastry, "Varieties of Learning Automata: An Overview," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 32, pp. 711–722, Dec. 2002.

[5] V. Gullapalli, *Reinforcement learning And Its Application on Control.* PhD thesis, Deparqement of Computer and Information Sciences, University of Massachusetts, Amherst, MA, USA, Feb. 1992.

[6] G. Santharam and P. S. S. M. A. L. Thathachar, "Continuous Action set Learning Automata for Stochastic Optimization," *Journal of Franklin Institute*, vol. 331B, no. 5, pp. 607–628, 1994.

[7] G. P. Frost, *Stochastic Optimization of Vehicle Suspension Control Systems Via Learning Auiomata.* PhD thesis, Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, Leicestershire, LE81 3TU, UI, Oct. 1998.

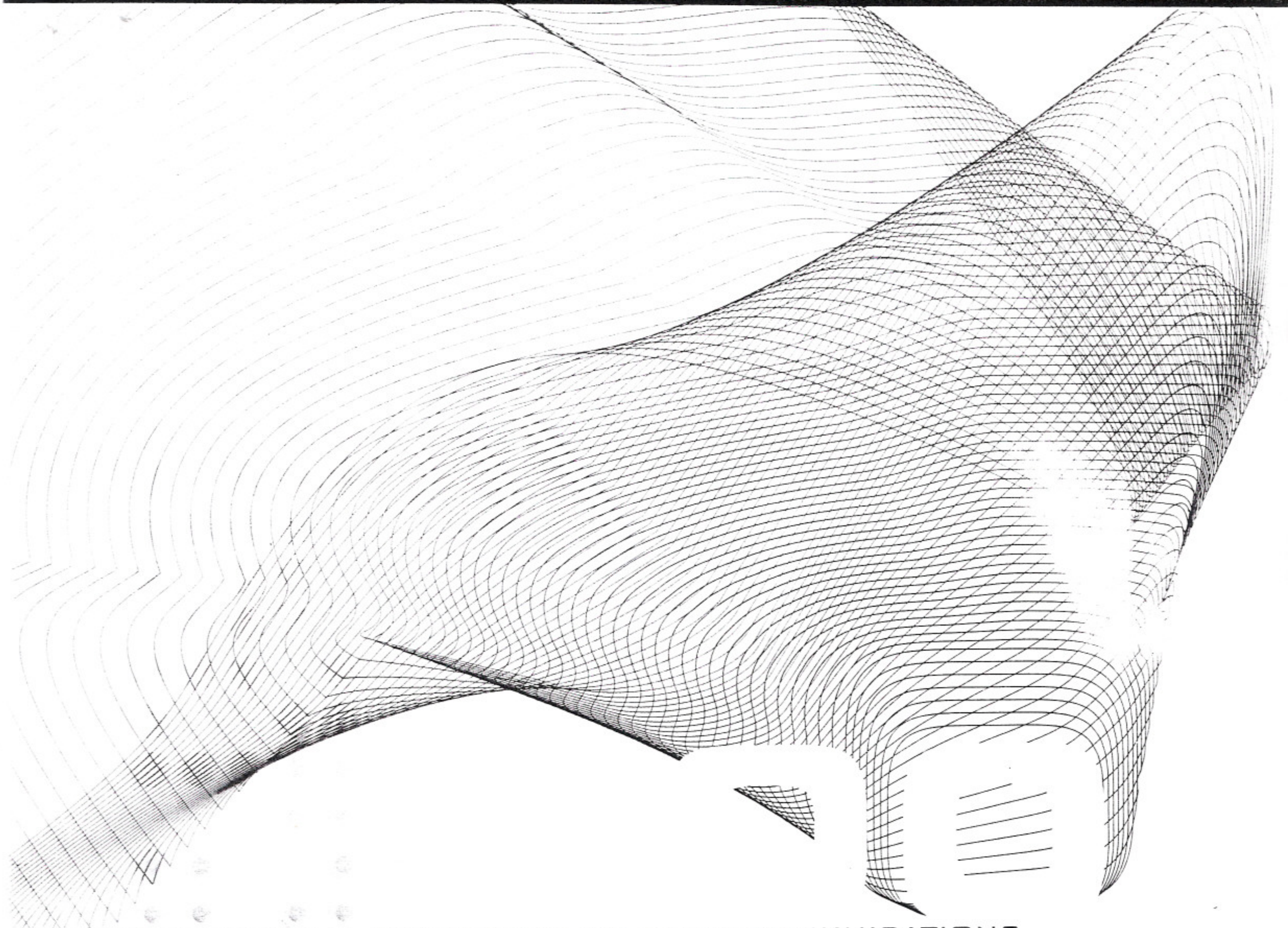
[8] G. I. Papadimitriou, S. M. S. Obaidat, and A. S. Pomportsis, "On the Use of Learning Automata in the Control of Broadcast Networks: A Methodology," *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*, vol. 32, pp. 815–820, Dec. 2002.

[9] B. J. Oommen and E. V. de St. Croix, "Graph Partitioning Using Learning Automata," *IEEE Transactions on Commputers*, vol. 45, pp. 195–208, Feb. 1996.

[10] M. R. Meybodi and H. Beigy, "A Note on Learning Automata Based Schemes for Adaptation of BP Parameters ," *Journal of Neuro Computing*, vol. 48, pp. 957–974, Nov. 2002.

[11] M. R. Meybodi and H. Beigy, "New Learning Automata Based Algorithms for Adaptation of Backpropagation Algorithm Parameters," *International Journal of Neural Systems*, vol. 12, pp. 45–68, Feb. 2002.

[12] H. Beigy and M. R. Meybodi, "A New Continuous Action-set Learning Automaton for Function Optimization," Tech. Rep. TR-CE-2003-002, Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran, 2003.

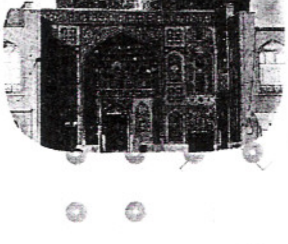


INTERNATIONAL SYMPOSIUM ON TELECOMMUNICATIONS

AUGUST 16-18, 2003
ISFAHAN,
IRAN

PROCEEDINGS

ITRC



DIALOGUE OF CIVILIZATIONS THROUGH TELECOMMUNICATIONS

IEE

IFIP TC6

KING'S College LONDON
University of London

IAEEE