Pattern-Recognizing Stochastic Learning Automata

ANDREW. G. BARTO, MEMBER, IEEE, AND P. ANANDAN

Abstract—A class of learning tasks is described that combines aspects of learning automaton tasks and supervised learning pattern-classification tasks. We call these tasks associative reinforcement learning tasks. An algorithm is presented, called the *associative reward–penalty*, or A_{R-P}, algorithm, for which a form of optimal performance is proved. This algorithm simultaneously generalizes a class of stochastic learning automata and a class of supervised learning pattern-classification methods related to the Robbins–Monro stochastic approximation procedure. The relevance of this hybrid algorithm is discussed with respect to the collective behavior of learning automata and the behavior of networks of pattern-classifying adaptive elements. Simulation results are presented that illustrate the associative reinforcement learning task and the performance of the A_{R-P} algorithm as compared with that of several existing algorithms.

I. Introduction

AMONG the many classes of tasks that have been formalized by theorists interested in learning are 1) the optimization problems under uncertainty focused upon by those studying learning automata and 2) the supervised learning pattern-classification tasks which have been studied extensively since the early 1960's. In this paper we describe a class of learning tasks that combines these two standard types of problems, and we present an algorithm for which we prove a form of optimal performance in these hybrid tasks by extending Lakshmivarahan's [1], [2] proof for time-varying stochastic automata. We call this algorithm the *associative reward–penalty*, or A_{R-P}, algorithm. Under one set of restrictions, it specializes to a stochastic learning automaton algorithm, in particular, to a nonlinear reward–penalty algorithm of the nonabsorbing type as characterized by Lakshmivarahan [2]. Under another set of restrictions, it specializes to the well-known perceptron algorithm of Rosenblatt [3] and, with a minor modification, to a supervised learning pattern-classification method based on the Robbins–Monro stochastic approximation procedure [4]. Our interest, however, is in the case in which both of these aspects of the algorithm operate simultaneously—yielding a pattern-classifying stochastic learning automaton.

This algorithm is the product of an effort to provide a formal basis for earlier results obtained by computer simulation [5]–[9]. Our interest in this algorithm is a result of our simulation experiments with networks of adaptive components implementing similar algorithms, and we discuss the implications of this approach below. Such adaptive elements constitute "self-interested" network components, the study of which was suggested to us by the theory of memory, learning, and intelligence developed by Klopf [10], [11] in which is proposed a class of algorithms similar to that studied here. To the best of our knowledge, the closest approximation to the particular algorithm presented here is the "selective bootstrap adaptation" algorithm of Widrow, Gupta, and Maitra [12], and we discuss this algorithm below. Also related is the much earlier stochastic pattern-classifying adaptive element of Farley and Clark [13]. As we discuss below (and as extensively discussed in [5] and [8]) later pattern-classification algorithms were designed for tasks that differ in an essential way from the task considered here. The A_{R-P} algorithm also has parallels with aspects of the "stimulus sampling theory" of mathematical psychology [14]–[17], although that research did not make contact with pattern-classification techniques. One study that combines stochastic automaton and pattern-classification algorithms is that of Jarvis [18], but the resulting algorithm is different from our own.

Before we define the learning tasks in which we are interested, which we call *associative reinforcement learning tasks*,[1] we briefly describe the most common learning automaton task, a commonly studied supervised learning pattern-classification task, and certain types of algorithms that have been applied to each.

II. Learning Automata

The theory of learning automata originated with the work of Tsetlin [19] and the independent research of mathematical psychologists [14]–[17], and has been extensively developed since then. Good reviews are provided by [20] and [21]. Also relevant is the independently developed line of research concerning the "two-armed bandit problem" (e.g., [22] and [23]). The framework in which learning automata have been most widely studied consists of an automaton and an environment connected in a feedback loop. The environment receives each action emitted by the

Manuscript received June 18, 1984; revised January 3, 1985. This work was supported by the Air Force Office of Scientific Research and the Avionics Laboratory (Air Force Wright Aeronautical Laboratories) through contract F33615-83-C-1078.

The authors are with the Department of Computer and Information Science, University of Massachusetts, Amherst, MA 01003, USA.

[1] In earlier publications by Barto and colleagues [5]–[9], these tasks were called *associative search tasks*, but we use the present terminology to avoid confusion with the unrelated usage of the same term in the field of information retrieval.