2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

Mutual Learning: Part I – Learning Automata

Kumpati S. Narendra
Center for Systems Science
Yale University
New Haven, CT 06521, USA
kumpati.narendra@yale.edu

Snehasis Mukhopadhyay
Computer and Information Science
Indiana University Purdue University Indianapolis
Indianapolis, IN 46202, USA
smukhopa@iupui.edu

***Abstract*—Learning theory has been studied for a long time by philosophers, and in the last century by psychologists and engineers. Yet, all learning is carried out in a general deterministic or stochastic environment, mostly by one isolated learner. This paper discusses the concept of mutual learning, where two or more entities attempt to learn from each other. The question posed is: "If two or more entities are learning in the same or similar environments trying to solve the same or similar tasks, how can they share their learning to improve themselves?" The authors believe that this is a central question that will keep researchers busy for many years. The paper merely introduces this question for discussion, and suggests some preliminary answers using the well known stochastic learning automaton framework for reinforcement learning.**

Keywords—mutual learning, reinforcement, learning automata, linear reward-inaction algorithm

I. Introduction

In the dictionary, the term "know" is defined as "learning or understanding gained through experience", and the term "learn" as "to gain knowledge through experience", underscoring the close relationship that exists between learning and knowledge. In their book "The Theories of Learning" (1981), Bower and Hilgard describe this relationship as being similar to that between a painting and a picture, and go on to say that if epistemology is the theory of knowledge, learning can aptly be called experimental epistemology.

1.1 Philosophy and Psychology

Philosophers have been analyzing the theories of knowledge for centuries, and more recently psychologists have investigated theories of learning since the beginning of the twentieth century. Formation of concepts, thoughts, and images, relationship between experiences and the organization of mind have all received a great deal of attention. These, in course of time, have merged with concepts in cognitive psychology, concerned primarily with the collection, transmission, storage and retrieval of information, and eventually led to the mathematical theories of Estes and Burke (1953), Bush and Mosteller (1955), Suppes and Atkinson (1960), and Hull (1963). The stimulus sampling theory of Estes has been the dominant approach within these mathematical theories.

All these indicate that contributions to learning have been vast, rich in ideas, and extend over many decades.

1.2 Engineering Systems

Since the ability of living organisms to cope with uncertainty is well known, it is only natural that efforts were made over the past century to incorporate similar features in engineering systems. This resulted in a variety of terms, borrowed from biology and psychology, to be introduced into the systems literature. These include adaptation, learning, and pattern recognition, as well as self-optimizing and self-organizing systems. Each one of these has developed over the years into an independent discipline with its own following. Adaptive control, which has been studied for over six decades, is concerned with the development of stable control laws for the adjustment of controller parameters in the presence of parametric uncertainty. Learning, which has a similar flavor, is concerned with improving the response of the system on the basis of past experience. Pattern recognition, developed mainly for the analysis of cognitive processes, is the method of classifying objects into predetermined classes. All these ideas are related to the broad, and at present very popular field of artificial intelligence (AI), which evolved from the disciplines of computer science, cognitive psychology, and automatic control (or, cybernetics), which refers to the machine emulation of higher mental functions. What is perhaps worth stressing is that, while the various subdivisions have flourished over the years, the border lines between them continue to be less than distinct, and the terms adaptive, learning, or AI systems can be used

978-1-5386-7926-5/$31.00 ©2019 AACC

interchangeably in many complex situations. Over the course of time, not surprisingly, advances made in one area spawned similar ideas in others, resulting in considerable overlap of basic concepts. In recent years, a new paradigm in adaptation, known as "Mutual Adaptation" was introduced by the first author and his co-workers. This paper is concerned with a similar concept in learning theory, referred to as "Mutual Learning".

1.3 Mutual Adaptation and Mutual Learning

In conventional adaptive control, the objective is to control a dynamic system in the presence of parametric uncertainty. The parameters of a suitable controller are adjusted adaptively, based on all available signals, so that the controlled system behaves in a desired fashion. This desired behavior is defined by a known, stable "reference model", chosen by the designer, which satisfies all the criteria set up by her/him. Achieving convergence of the behavior of the dynamics of plant together with the adaptive controller to that of the reference model, in a stable and robust fashion, has been the objective of adaptive control over the past sixty years.

During the past six years, the first author and his co-workers have been working on problems in which two (or more) dynamical systems adapt to each other. The importance of such problems arises from the fact that in many practical applications, no simple reference models can be used, and each system attempts to improve its behavior based on observations made on both its own response as well as the responses of the others. This has been defined as "Mutual Adaptation". To improve their responses, the two systems use each other as their own reference models implicitly. Assuring the stability of such mutual adaptation consequently poses a major problem. Deriving sufficient conditions for stability, using ideas borrowed from conventional adaptive control, proved unsuccessful over a period of six years, even for two simple dynamical systems of second order. It was only recently that a new approach to mutual adaptation was proposed by Narendra and Phillips that assured stability in all simple cases investigated. The principal idea behind this new approach is that the two systems should asymptotically adapt to each other alternately (i.e., only one system adapts at any moment of time). Work is currently in progress to extend this concept to more complex dynamical systems. The authors of this paper are interested in both adaptive control and learning control, and it was only natural that they should attempt to extend the concept of mutual adaptation to learning systems as well. This is what is termed "Mutual Learning". Though the two concepts at first seem similar, a deeper investigation reveals that they lead to large and complex questions that are vastly different. Our principal objective in this paper is to introduce this important concept of "Mutual Learning".

1.4 Learning Theory Today

In their book "Reinforcement Learning: An Introduction", Sutton and Barto (1998) state that the history of reinforcement learning consists of two threads, both of which have evolved over a long period and both of which enjoy a rich literature. The first is learning by trial and error, which started in psychology to explain behavior patterns in living systems, which were adapted to engineering systems as "learning automata" [] in the 1960s. This led to the revival of reinforcement learning in the 1980s. The other thread is optimal control, which is an integral part of control theory. This does not strictly involve learning, but is based on the pioneering work of Pontryagin and his co-workers, as well as the classical theories of Hamilton and Jacobi, and the Principle of Optimality of Bellman. Today learning theory is a vast field, multidisciplinary in character in which a large number of very different models are being investigated. The theory is finding applications in such diverse areas as machine learning, multi-agent systems, and neuroscience. While the models were initially static, the approaches developed were extended to dynamic Markov Decision Processes (MDP) with finite states, and later to stochastic nonlinear difference equation models. All the principal learning models suggested in the literature fall within the scope of the investigations proposed in this paper.

1.5 Some Related Research

Ikemoto et al (2012) discuss an interesting study involving a human-robot mutual learning and co-adaptation, inspired by the parenting behavior in humans. In the context of artificial neural networks, Zhang et al (2017) discuss the problem of an ensemble of deep neural networks learning from each other in the context of a classification task. It was concluded that a collection of small neural networks with mutual learning can outperform a "powerful" single teacher network. In the context of machine vision, Nie et al (2018) discuss mutual learning to achieve superior performance between two related, but disparate computer vision tasks, i.e., human parsing and pose estimation. Another related research theme is multi-agent learning systems, where the agents focusing on different, disparate subtasks of a complex task cooperate to solve the problem, in a spirit similar to mathematical game theory. Panait and Luke (2005) provides a survey of this somewhat well-established field, highlighting the issues of inter-agent communication, task decomposition, and scalability in such multi-agent systems. In contrast to the earlier theme of multi-agent systems, in mutual learning, the agents are involved in solving the same, or similar, tasks, and the objective is for them to act as (partial) teachers to each other in order to improve their learning. Further, our objective, in contrast to other works, is to study such mutual learning problem as a systems theory problem, focusing on general questions and issues. We also, for the first time, use the learning automata paradigm (Narendra and Thathachar, 2012) (an early, but provably convergent reinforcement learning approach for static

environments) in order to study mutual learning. The mathematical convergence properties of such learning automata algorithms are well established, and our interest is in determining the changes that have to be made in the theory to accommodate mutual learning.

II. The Objectives and The Scope of the Paper

Learning is a loose open concept that includes diverse areas. Ideas of learning have arisen from different scientific communities such as reinforcement learning, robotics, optimal control, dynamic programming, and complex systems. Naturally, many of the questions raised have been addressed using the languages of the corresponding communities. The authors of this paper are systems theorists, interested in both learning and control theories, and in formulating problems using a mathematical framework.

As stated in Section 1.3, our principal objective in this paper is to introduce the important concept of "Mutual Learning" and to initiate discussions on the topic among members of the systems community. The authors believe that problems involving mutual learning are ubiquitous, and range from relatively simple to those that are extremely complex and difficult to formulate analytically.

Assuming that two or more agents have "learned" partially about a process, an environment, or a specific situation, the principal question to be addressed is how they should share their learning to improve themselves. In practice, the learning procedures adopted may be identical, similar, or vastly different, and these may, in turn, render mutual learning increasingly complex and difficult to describe analytically. Our principal aim in this paper is to discuss some of the questions that arise when two learning automata (algorithms) operating in a random environment attempt to learn from each other.

Mutual learning can happen between two humans, a human and a machine, or two machines. The first two classes are of direct interest even to researchers outside of the systems community such as social psychology. However, mutual learning between machines will be of greater interest in technology, an interest that is bound to grow with advances in autonomous systems. Assuming that the two machines are learning from each other about the same or similar processes, our primary interest will be in the precise mathematical formulation of the problems that arise – to the extent possible.

It is well known that various methods such as supervised learning, unsupervised learning, and reward based learning have been addressed in the learning literature. Depending upon the problem addressed, any of these approaches can be used in the formulation of the above problems.

III. Mutual Learning Scenarios

As stated in the introduction, the concept of "mutual learning" arises in infinitely many situations, which are hard to classify. However for convenience, as well as the fact that they lend themselves to quantitative analysis, we consider in this section 7 different scenarios as follows.

Mutual Learning of the Optimal Action: The first scenario, which deals with learning automata, is the main subject of this paper, and is considered in some detail in later sections.

Mutual Learning in a Classification Task: The judgment that some part of the experience belongs to a specific class is "pattern recognition". Examples of such pattern recognition include: (i) recognition of a tune, (ii) recognition of a friend's house, (iii) recognition of a rare wine by its taste, and (iv) recognition by an instructor whether or not a course is enjoyed by the students. As an illustration of mutual learning, if two persons or machines recognize a tune differently, how should they proceed, on the basis of the information received from the other?

Mutual Learning in Identification: Two different algorithms are used to identify the parameters of an unknown system. The output error of the first algorithm is smaller (according to some metric) than that of the second. What actions should the two algorithms take to make use of each other's information?

Mutual Learning in a Static or Dynamic Optimization Task: Two identical algorithms, but with different initial conditions, attempt to maximize a performance function. After a finite number of steps, what information should they exchange to improve their convergence?

Mutual Learning of the graph topology in an uncertain and/or changing network decision problem: Two crawling agents are exploring and learning two subgraphs of an overall large graph. How should they exchange information concerning their respective subgraphs so that a decision problem in the overall large graph can be solved efficiently?

Mutual Learning to add new consistent rules or eliminate existing inconsistent ones in a rule-based system: Two rule-based agents have mutually complementary logical rules. Each is consistent and correct by itself; however, if they combine the rules, inconsistencies may arise. How do they learn consistent, correct rules from each other?

Mutual Learning of unknown terrains and/or state space in a cooperative terrain or target acquisition problem: Consider the problem of multiple robots cooperatively developing the map of an unknown, possibly hazardous terrain. In order to combine their knowledge, landmarks in their respective terrain maps need to be registered and made to correspond to each other. How do they learn to register each other's information?

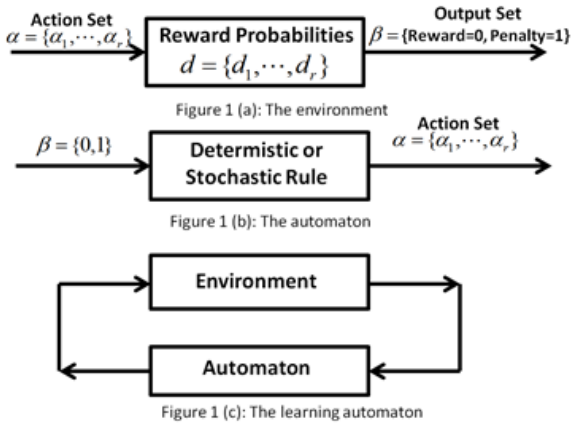
IV. MUTUAL LEARNING OF OPTIMAL ACTIONS IN STOCHASTIC LEARNING AUTOMATA

As stated in the introduction, learning theory is a vast field which is multidisciplinary in character. Our objective in this paper is to introduce the concept of mutual learning in a simple context, and provide a sense of the multitude of questions it can give rise to.

In the previous section, seven scenarios where mutual learning arises were presented. The first of these, the main subject of this paper, deals with trial and error learning investigated over the past six decades, and referred to as learning automata. The other six scenarios will be considered in future papers. The authors believe that the questions raised in this paper will also be relevant in future contributions.

a) The Learning Automaton:

A learning automaton consists of an environment E connected in feedback with an automaton A as shown in Figure 1 (c).



An environment E (figure 1a) is described by the triple $\{\alpha, d, \beta\}$ where $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ represents the finite input set with r actions α_i, an output set $\beta = \{\beta_1, \beta_2\}$ with $\beta_1 = 1$ representing a reward, and $\beta_2 = 0$ representing a penalty, and $d = \{d_1, d_2, \ldots, d_r\}$ a set of unknown reward probabilities $d_i = \text{Prob}[\beta(n) = 1 | \alpha(n) = \alpha_i]$.

An automaton A (figure 1b) takes in a sequence of inputs which are the outputs of the environment, and puts out a sequence of actions as inputs to environment using a deterministic or stochastic rule. The objective is to "learn" the environment's responses to different actions and gradually evolve so that actions with better environmental responses are chosen with higher probabilities. Such an automaton which improves its actions while operating in a random environment is called a "learning automaton"(figure 1c). If, in the limit, only the best action (the one with the highest reward probability) is chosen with probability 1, the automaton is said to be "optimal". If the response of the automaton in the limit is better than when all actions are chosen with equal probabilities, the automaton is said to be "expedient". Our interest is in designing learning automata which are arbitrarily close to optimality.

Deterministic and Stochastic Automata: If the rule by which the automaton makes future choices about actions is deterministic, the automaton is called a deterministic automaton. If the automaton's action probabilities are updated after every event, and the action is chosen probabilistically by sampling the action probabilities, the automaton is said to be stochastic.

Comment: In this paper we consider mutual learning when stochastic learning is used. Future papers will discuss interactions between both deterministic and stochastic schemes.

Comment: A very large number of both deterministic and stochastic automata with very different convergence properties, have been reported in the literature. After a certain number N of trials, the automaton has obtained some information about the environment. Our interest is in the manner in which an automaton can use information provided by the others to improve itself.

Even in this very simple learning scenario, a large number of questions in mutual learning can be posed. Below, we merely list 4 of these questions, the answers to which reveal the nature of the difficulties encountered in mutual learning:

Question 1: Assuming that two stochastic automata with two actions using *identical* learning algorithm, but performing n_1 and n_2 trials respectively result in actions probabilities $(p_1(n), p_2(n))$ and $(q_1(n), q_2(n))$ respectively, how should they change their probabilities based on those of the other to improve their performance?

Question 2: How would the above procedure change, if the two learning automata have different learning parameters (e.g., the learning step sizes are different)?

Question 3: Another more complex situation is where one of the automata is optimal, while the other is only expedient.

Question 4: In all the above questions, it was assumed that number of actions for each automaton is only 2. How would these procedures change if the number of actions is greater than 2.

b) Mutual Learning in Stochastic Learning Automata:

In stochastic learning automata, an action probability vector $p(n)$ is associated with the action set α. $p^T(n) = [p_1(n), p_2(n), \ldots, p_r(n)]$ where $p_i(n)$ represents the probability with which action α_i is selected at trial n. At every stage, using $p(n)$ an action is selected an performed in the environment. Based on the response $\beta(n)$ of the environment, $p(n)$ is updated as $p(n+1)$ as follows:

$$p(n+1) = T[p(n), \alpha(n), \beta(n)] \qquad (1)$$

Equation (1) in essence represents the essence of "learning" that takes place at every trial. While a very large number of linear and nonlinear reinforcement schemes have been reported in the literature (Narendra and Thathachar, 2012), we consider in this section only a few of them for discussion.

The L_{R-I} Algorithm (2 actions): The automaton has initial probabilities $p_1(0) = p_2(0) = 0.5$, chooses one of the actions α_i with these probabilities, and performs that action. If the action results in a reward, it increases the probability $p_i(n)$ and decreases the probability of the other action. If the result is a penalty, the action probabilities are left unchanged, and the experiment is repeated. More precisely,

$$p_i(n+1) = p_i(n) + a\big(1 - p_i(n)\big),$$
$$\text{if } \alpha(n) = \alpha_i, \beta(n) = \text{reward}$$
$$p_i(n+1) = p_i(n), \text{otherwise}$$

The probability of the other action is adjusted so that the two action probabilities add up to 1.

Comment: It is seen that a decrease in an action probability occurs only when the other action results in a reward.

The algorithm is referred to as the Linear Reward-Inaction algorithm, and has been proved to be ϵ-optimal.The first experiment below is concerned with such automata.

Experiment 1: An environment has two input actions α_1 and α_2, and the corresponding reward probabilities are $d_1 = 0.5$, and $d_2 = 0.05$. This implies that the first action is superior to the second.

We consider the two automata are conducting independent experiments on the environment. The convergence of the two shown in Figure 2 for different step sizes of 0.01, 0.05, 0.1, and 0.2. In the experiment that follows a step size of 0.01 is chosen. If one of the automata is allowed to operate more than 400 steps, it becomes evident that the first action is better than the second one. If the other automaton has not started yet, then it can be considered as a "student" learning from the first automaton, "the teacher", that action 1 is superior.

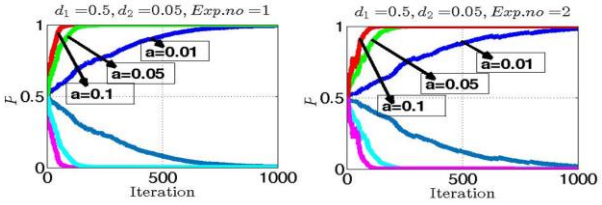


Figure 2: Two action automaton
(step sizes: 0.01, 0.05, 0.1)

Experiment 2: We next consider the case when A_1 has conducted experiments 300 times, and A_2 has conducted experiments only 100 times. The issue of mutual learning arises in this case. How do the automata change their probabilities based on the information provided by the other?

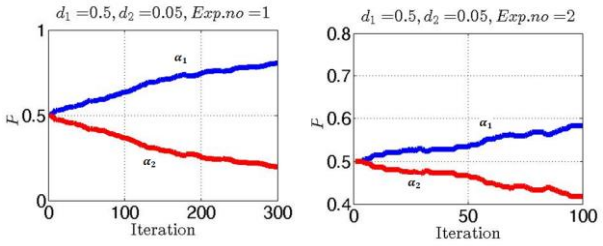


Figure 3: Experiment 2

There are a variety of answers that can be provided for this question that we shall consider elsewhere. If the two automata have tried action 1 N_1 and $\overline{N_1}$ times, and received rewards M_1 and $\overline{M_1}$ times, a simple decision rule may be to assume the unknown reward probability to be $(M_1 + \overline{M_1})/(N_1 + \overline{N_1})$. Similarly, the reward probability of the other action can also be determined.

Experiment 3: The evolution of the action probabilities when $d_1 = 0.5$ and $d_2 = 0.35$ are shown in Figure 4. Once again, the same questions arise in this problem, but since the reward probabilities are close to each other, the exchange of information is more critical. It depends upon the detailed information that one automaton can provide the other about their trials. If all information is assumed to be shared about the trials, both of them can conduct virtual experiments based on the other's trials. This is shown in Figure 4.

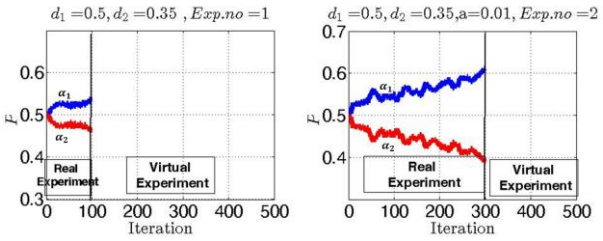


Figure 4: Experiment 3: Real and Virtual Experiments by the Two Automata

Comment: In the experiments thus far, we have considered two automata using the same algorithm with the same step size. In practice, this is rarely the case. One automaton may be using an L_{R-I} scheme, while the other may be using an L_{R-P} or $L_{R-\epsilon P}$ with a different step size. How the two automata should alter their strategies is the key question.

In the next two experiments we consider a stationary environment with an action set consisting of five actions $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$. In experiment 4, the two automata perform 100 and 500 trials using all five actions respectively. Note that the best action at this stage for Automaton A1 is $\propto 4$ and for Automaton A2 it is $\propto 1$.

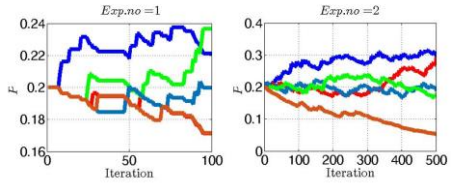


Figure 5: Experiment 4: Automata with 5 Actions

The question once again is the best decision that the two automata can make based on their interactions. In experiment 5, the first automaton tries only the first two actions, while the second automaton tries the last three actions. The evolution of the action probabilities are shown in Figure 6. Based on the exchange of information, how can they conclude that $\propto 1$ is the best action in the entire set?

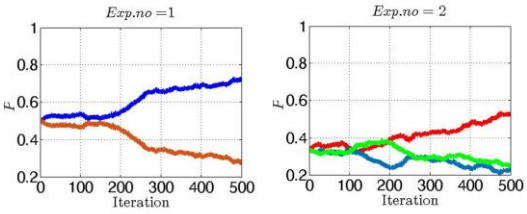


Figure 6: Experiment 5

V. COMMENTS AND CONCLUSIONS

The paper introduces the concept of "mutual learning" in which two (or more) agents which "learn" in a random environment, attempt to improve themselves based on the information provided by the other(s). In a sense the concept is not new, since "divide and conquer" strategies have been used to solve many complex problems in the past. The essential difference between what is proposed in the current paper and past work is the suggestion that the problems should be formulated within a mathematical framework and that the changes in each participant should be quantified in some fashion.

The last section of the paper deals with two stochastic learning automata operating in a stationary random environment. Even in this relatively simple case, only general comments can be made at this stage of research; substantive and definitive answers require more work by the research community. The authors believe that the paper will give rise to interesting and meaningful discussions in the systems community concerning mutual learning.

ACKNOWLEDGMENT

The authors are very grateful to Kasra Esfandiari for his help with the excellent simulation studies included in this paper.

REFERENCES

[1] Bower, G.H. and Hilgard, E.R., 1981. *Theories of learning*. Prentice-Hall.

[2] Bush, R. R. and Mosteller, F. Stochastic models]or learning. New York: Wiley, 1955

[3] Estes, W.K. and Burke, C.J., 1953. A theory of stimulus variability in learning. *Psychological Review*, *60*(4), p.276.

[4] Hull, C.L., 1963. 13 THE UNIFORMITY POINT OF VIEW1. *Theories in Contemporary Psychology*, p.203.

[5] Ikemoto, S., Amor, H.B., Minato, T., Jung, B. and Ishiguro, H., 2012. Physical human-robot interaction: Mutual learning and adaptation. *IEEE robotics & automation magazine*, *19*(4), pp.24-35.

[6] Narendra, K.S. and Thathachar, M.A., 2012. *Learning automata: an introduction*. Courier Corporation.

[7] Nie, X., Feng, J. and Yan, S., 2018. Mutual Learning to Adapt for Joint Human Parsing and Pose Estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 502-517).

[8] Suppes, P. and Atkinson, R.C., 1960. Markov learning models for multiperson interactions Stanford. *Press, i960*.

[9] Sutton, R.S. and Barto, A.G., 1998. *Reinforcement learning: An introduction*. MIT press.

[10] Zhang, Y., Xiang, T., Hospedales, T.M. and Lu, H., 2017. Deep mutual learning. *arXiv preprint arXiv:1706.0038.*