MODELING, IDENTIFICATION, AND CONTROL OF LARGE-SCALE DYNAMICAL SYSTEMS

Simon Haykin, McMaster University (haykin@mcmaster.ca)
McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1
Alfred O. Hero III (hero@eecs.umich.edu)
University of Michigan,1301 Beal Avenue, Ann Arbor, MI 48109-2122
Eric Moulines (moulines@tsi.enst.fr)
ENST, 46, rue Barrault, 75634 PARIS Cédex 13

ABSTRACT

This paper highlights some fundamental issues involved in the study of large-scale dynamical systems. Two particular topics are discussed in some detail, one dealing with the management of active sensors via partially observable Markov decision processes, and the other dealing with the modeling, recognition and tracking of multi-function radars in an electronic warfare environment.

I. INTRODUCTION

All dynamical systems share a basic feature: the *state* of the system, be it scalar or vector, varies with *time*. Typically, the state is not measurable directly. Rather, in an indirect manner, it makes its effect measurable through a set of *observables*. As such, the characterization of dynamical systems is described by a *state-space model*, which, in general, embodies two equations:

(i) *State-evolution equation*, which describes the evolution of the state as a function of time:

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \mathbf{w}_t \tag{1}$$

where t denotes discrete time, $\mathbf{x}_t$ denotes the state vector at time t, $\mathbf{f}(.)$ is a vector-valued function of its argument, and the vector $\mathbf{w}_t$ denotes *dynamic noise*.

(ii) *Measurement equation*, which takes one of two forms, depending on whether the system is passive or active:

(a) *Passive dynamical system*, described by

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t) + \mathbf{v}_t \tag{2a}$$

where the vector $\mathbf{y}_t$ denotes the set of observables, $\mathbf{g}(.)$ denotes another vector-valued function, and the vector $\mathbf{v}_t$ denotes *measurement noise*.

(b) *Active dynamical system*, described by

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t, \mathbf{a}_t) + \mathbf{v}_t \tag{2b}$$

where the additional vector $\mathbf{a}_t$ denotes *action* taken by the system at time t.

According to (2a) and (2b), it is the action $\mathbf{a}_t$ that distinguishes an active dynamical system from a passive one. Most important, an active dynamical system *explores* its environment by taking action $\mathbf{a}_t$ whenever the environment resides in state $\mathbf{x}_t$; we may therefore think of $(\mathbf{x}_t,\mathbf{a}_t)$ as a *state-action pair*. Just as the environment state $\mathbf{x}_t$ spans a *state space*, the action $\mathbf{a}_t$ spans a space of its own called the *action space*. The constituents of the action space may be different modalities, waveforms, functions, etc., over which the system is able to operate. On this basis, we say an active dynamical system is of a *large-scale* kind due to a combination of three factors:

(i) high dimensionality of the environment state space;
(ii) high computational complexity of the nonlinear predictive model used in tracking the state of the environment; and
(iii) high search complexity of the action space.

In contrast, a passive dynamical system merely *listens* to its environment; and through observables produced by the environment, it infers the state of the environment. Accordingly, a passive dynamical system is said to be of a *large-scale* kind solely on the basis of factors (i) and (ii).

The availability of an action space or its absence has serious implications for the specific functions which a dynamical system can perform. Specifically, an active dynamical system is capable of interacting with its environment; hence, through searching over the action space for an optimal policy, it has a natural capability to perform *optimal control*. On the other hand, a passive dynamical system is well positioned to *model* its environment and use the model for the purposes of *classifying* the environment and *tracking* its state.

Section II of the paper discusses the optimal scheduling policy in active sensor networks consisting of a large number of active sensors (not necessarily of the same type); the policy involves the mapping from past observations to future actions. Section III discusses the signal-processing issues that arise in the passive modeling of multi-function radars in an electronic warfare environment; the issues of recognition and tracking of such radars are briefly mentioned. Section IV concludes the paper.

II. MANAGEMENT OF ACTIVE SENSOR NETWORKS VIA PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES

In active sensing, we sequentially and adaptively schedule sensors, modalities, waveforms, or search patterns as a function of past measurements. The theory of partially *observable Markov decision processes* (POMDP) is used to determine an *optimal scheduling policy* (i.e., a mapping from past measurements to a future action) from the statistical model of the sensor and the environment. Such a policy is used by the sensor management algorithm. The unconstrained globally optimal policy is to deploy all sensors, modes, waveforms, and search patterns simultaneously but this is impractical since resources such as energy, computation and deployment agility are always limited. In general, active sensing must account for deployment constraints and balance complex tradeoffs between competing mission goals such as detection of new targets, tracking and identification of existing targets.

The value of a particular scheduling policy relative to a specific goal is captured by a *reward* function that depends on the true state of target in addition to the goal and policy. The objective of active sensing in the POMDP framework is to find policies that will maximize the average expected reward over time. This leads to optimal sensor-scheduling policies that depend on the posterior distribution of the system state conditioned on sensor measurements. In tracking applications, the system state describes probabilistically *uncertainty* in the number of targets, locations of the individual targets, and movements of the targets. It may also describe uncertainty in sensor characteristics (e.g., spatial position, or clutter).

Sensor-scheduling strategies may be *myopic*, involving single-stage, one-step prediction, or *non-myopic*, involving multi-stage, multi-step prediction. In the myopic case, sensing actions are taken so as to maximize the immediate reward and do not attempt to predict the effect of an immediate action on the more distant future. Optimal sensor-scheduling policies are almost always non-myopic, since each action must maximize all future rewards and therefore such policies must predict the future value of information gained from each action. While myopic methods have the advantage that they are more computationally tractable than non-myopic methods, they are usually suboptimal. Several researchers have developed approximate solution techniques for optimal non-myopic sensor scheduling (e.g., the general rollout algorithms of [3], the multi-arm bandit active beam scheduling approximations of [8], the value-to-go and reinforcement learning (RL) approximations developed in [5], [6], [7] for active sensing). In the RL approach, described in more detail below, training examples of sensing actions, responses, and the observed system states are used to learn an optimal sensor-scheduling policy.

A. POMDP Framework

A discounted-reward Markov decision process (MDP) is defined by a Markovian sequence of states $\{\mathbf{x}_t\}_{t\geq 0}$ taking values in a state space χ, a sequence of causal actions $\{\alpha_t\}_{t\geq 0}$ taking values in an action space A, and a (possibly random) reward function $R_t = r_t(\mathbf{x}_t,\alpha_t)$ that assigns the cost incurred (when negative) or the reward gained (when positive) to the event of being at state $\mathbf{x}_t$ and taking action α_t. The state space can contain rich information such as the number of targets present, their locations, their type, and whether they are fixed or moving. Each action is a causal function of the state sequence and specifies which sensor to use, the mode of operation, and where to point the sensor. The reward system reflects the tradeoffs between costs of deploying a certain sensor mode and the gain earned from the measurements it collects. Consider the scenario where one observes the current state $\mathbf{x}_t$, chooses action α_t and observes the state transition to $\mathbf{x}_{t+1}$; generating the state-action sequence $\mathbf{x}_1,\alpha_1;\mathbf{x}_2;\alpha_2;\ldots$. This sequence is called an MDP, since, given $\mathbf{x}_t$ and α_t, $\mathbf{x}_{t+1}$ is independent of all past states and actions. When only an indirect measurement $\mathbf{y}_t$ of the state $\mathbf{x}_t$ is available, the action sequence must be based on $\mathbf{y}_t$ rather than $\mathbf{x}_t$, and the posterior density $p(\mathbf{x}_t|\mathbf{y}_t,\alpha_t;\ldots;\mathbf{y}_0,\alpha_0)$ carries all relevant causal information about the state. For this reason, $\tilde{\mathbf{x}}_t = p(\mathbf{x}_t|\mathbf{y}_t, \alpha_t;\ldots;\mathbf{y}_0, \alpha_0)$ is called the *information state*. The state-action sequence $\tilde{\mathbf{x}}_1, \alpha_1;\tilde{\mathbf{x}}_2, \alpha_2;\ldots$ defines a POMDP. As the theory of POMDP parallels that of MDPs, we suppress the "tilde" on $\mathbf{x}$ in the sequel.

A stationary policy Π is a map from the state space χ to the action space A that specifies the action taken at each state. Denote the class of all policies by P. The value function associated with policy Π, denoted by $V^{\Pi}(\mathbf{x})$ is the expected total discounted reward when in state $\mathbf{x}_t=\mathbf{x}$ and following policy Π, that is

$$V^{\Pi}(\mathbf{x}) = E\left\{\sum_{\tau=t}^{\infty}\beta^{\tau-t}r(\mathbf{x}_\tau, \Pi(\mathbf{x}_t))\middle|\mathbf{x}_t=\mathbf{x}\right\} \quad \forall\mathbf{x}\in\chi \qquad (3)$$

where $\beta\in(0,1)$ is a *discount factor*, included to reduce the value of future rewards as compared with immediate rewards. The conditional expectation is taken with respect

to the joint distribution of all the targets. An optimal policy maximizes the value function for all t, which is defined by the unique solution to Bellman's equation:

$$V(\mathbf{x})=\max_{\alpha} E\{r(\mathbf{x}_t, \alpha)+\beta V(\mathbf{x}_{t+1}) \,|\, \mathbf{x}_t\} \;=\; \mathbf{x}, \alpha_t=\alpha \qquad (4)$$

Unfortunately, when cardinality of the state and action spaces are large and the state transition density is either computationally complicated or not explicitly available, Bellman's equation becomes computationally intractable.

B. Q-Learning with Function Approximation

Define the function $Q(\mathbf{x},\alpha)$ as the conditional expectation on the right-hand side of (4). Q-learning is a special case of RL using a particular type of stochastic approximation to approximate the Q-function and extract a near-optimal policy via $\alpha_t = \max_\alpha Q(\mathbf{x}_t,\alpha)$, t = 1,2,.... For POMDPs, Q-learning cannot be applied directly due to the continuity of the information state space, which makes the Q-function infinite-dimensional. Therefore, a finite-dimensional function approximation is used to approximate the Q-function, for example, using a with truncated linear basis expansion

$$Q(\mathbf{x}, \alpha) \approx \theta^T \phi(\mathbf{x}, \alpha) \qquad (5)$$

where ϕ is a vector of known basis functions and $\theta \in \Re^L$ is an unknown coefficient vector to be approximated.

In batch Q-learning, Q is estimated from a set of simulated state-action sequences. Specifically, the training process involves the generation of *state, action, next state, immediate reward* 4-tuples over a large number of training episodes. This set of training episodes is used in batch to estimate the Q-function for a particular state-action pair. Given an arbitrary initial value of $Q_k(\mathbf{x},\alpha)$ at iteration k = 0, the one-step Q-learning algorithm is given by repeated application of the update equation

$$Q_k(\mathbf{x}, \alpha)=(1-\gamma)Q_{k-1}(\mathbf{x}, \alpha)+\gamma\left(r+\beta\max_{\alpha \in A} Q_{k-1}(\mathbf{x}', \alpha)\right) \qquad (6)$$

where each of the 4-tuples $\{x_t=\mathbf{x}, \alpha_t=\alpha, \mathbf{x}_{t+1}=\mathbf{x}', R_t=r\}$ are incurred during the progress of the MDP, and the discount factor $\gamma \in (0, 1)$ decreases with t. When γ decreases to zero as $a/(b+t)$ where $(a,b>0)$, this algorithm converges to the true Q-function with probability 1, regardless of the actual policy used in generating the trajectories as long as the state-action pairs are visited infinitely often [2]. With function approximation, a simple gradient-descent method is commonly used to update the estimate $\hat{\theta}$ of θ:

$$\hat{\theta}_k=\hat{\theta}_{k-1}+\gamma\left(r+\beta\max_{\alpha}\hat{\theta}_{k-1}^T[\phi(\mathbf{x}', a')-\phi(\mathbf{x}, a)]\right)\phi(\mathbf{x}, a)$$

For more details regarding the implementation of Q-learning with function approximation for active-sensor management, the reader is referred to [5], [6], [7].

III. MODELING, RECOGNITION, AND TRACKING OF MULTI-FUNCTION RADARS

The theory of partially observable Markov decision processes (POMDP) and related issues also arise in the study of *electronic support* (ES), which is a field of electronic warfare. The function of an ES system is to infer the state of an *uncooperative* radar system in a battlefield environment. This function is achieved by *passively* sensing and operating on the sequence of electromagnetic pulses emitted by the radar. To assist this critical function, the ES system also exploits prior knowledge gathered about the radar system that could be operating in the battlefield.

Much has been written on the study of ES systems pertaining to previous generations of mechanically scanned antennas [12]. However, when confronted with the new generation of *multi-function radars* (MFRs), the ES analysis of signals emitted by such radars is significantly more complex due to two realities:

(i) Phased array antennas are used to electronically scan the radar environment in a highly agile manner.
(ii) Flexible sophisticated software control algorithms are used to perform radar functions (i.e., searching, acquisition, and tracking) on multiple targets, virtually simultaneously.

Basic to the ES analysis of signals emitted by an MFR is the development of a *model* that can facilitate the joint tasks of MFR recognition and state estimation. To that end, the MFR is viewed as a *discrete-event system*, which speaks some known, or partially known, formal language [4]. Correspondingly, the sequence of observations of MFR signals is viewed as strings from this language, corrupted by measurement noise. In [9], [10], [11], it is shown that by using prior knowledge about some particular MFR signal, it is indeed possible to generate a grammar that describes the radar language. Such an approach to the modeling of MFRs is referred to as *syntatic modeling*. The important point to note here is that syntatic models are compact formal representations that can form a homogeneous basis for modeling the complex dynamics being performed by MFRs. The adoption of syntatic models for MFRs as proposed in [9], [10], [11], represents the basis of a *model-centric approach* to the

design of ES systems, with the syntatic model being viewed as a *compressor* of data in the electronic warfare library pertaining to the MFR in question. This modern approach to the design of ES systems should be contrasted with the *data-centric approach* adopted in the traditional electronic warfare literature.

Given the syntatic models of possible MFRs operating in a battlefield environment, we may now envision the design of an ES system whose function is to perform the following pair of functions in real time:

(i) *Recognition*, the purpose of which is to infer the particular type of MFR that is responsible for emitting the observed sequence of electromagnetic pulses.

(ii) *State estimation*, the purpose of which is to infer the state (i.e., searching, acquisition, or tracking mode) in which the MFR is operating.

These two functions are inter-related and must therefore be performed *jointly.* In point (ii), note also that the MFR state is not interesting; rather, what is important is the mapping of MFR state into an *instantaneous threat*.[1]

In their own respective ways, MFR recognition and state estimation involve making decisions in the face of *uncertainty.* The sources of uncertainty include measurement noise, ES system imperfections, and incomplete knowledge about the MFR that would be operating in the battlefield environment. This incomplete knowledge may be viewed as an *information gap*. In any event, given the serious consequences of the decision-making process, optimality of the ES system may have to be sacrificed in favor of *robustness*. In this context, the strategy referred to as "information-gap decision theory: decisions under severe uncertainty" in [1] deserves particular attention.

There is another issue that needs to be considered in the design of ES systems, namely, the *allocation of resources*. Recognizing that computing resources are limited and faced with the possibility of having to deal with more than one MFR, we have the setting for an additional requirement: which particular MFR is likely to pose the greatest threat and therefore warrants the focusing of limited resources? This problem becomes even more challenging when the MFRs form a *cooperative network* designed to share information between themselves.

1. On a related note, ES has a complementary field known as *electronic attack*, where some action (e.g., jamming) is taken to defeat the threat.

IV. CONCLUSION

The study of large-scale dynamical systems is emerging as one of the fields likely to dominate the twenty-first century. This field of study is multidisciplinary in nature, permeating many areas:

- Engineering, exemplified by signal processing, control, communications, computers and biomedical;
- Physical sciences, exemplified by geophysics, and nanoscience;
- Biological sciences, exemplified by neuroscience, and neural-information processing systems;
- Economics.

In this introductory paper, we have highlighted some fundamental issues that arise in the study of active and passive dynamical systems that are of a large-scale kind.

REFERENCES

[1] Y. Ben-Haim, *Information-gap Decision Theory: Decisions under Severe Uncertainty*, San Diego: Academic Press, 2001.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

[3] D. P. Bertsekas and D. Castanon, "Rollout Algorithms for Stochastic Scheduling Problems", *Journal of Heuristics*, Vol. 5, pp. 89-108, 1999.

[4] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Boston, MA: Kluwer Academic, 1999.

[5] C. Kreucher, A. O. Hero, D. Blatt, and K. Kastella, "Adaptive multi-modality sensor scheduling for detection and tracking of smart targets," in W*orkshop on Defense Applications of Signal Processing*, 2004.

[6] C. Kreucher, A. O. Hero, K. Kastella, and D. Chang, "Efficient methods of non-myopic sensor management for multitarget tracking", *IEEE Conf. on Decision and Control*, Bahamas, 2004.

[7] C.Kreucher and A.O. Hero, "Non-myopic approaches to scheduling agile sensors for multistage detection, tracking and identification," *ICASSP-2005*, Philadelphia, March 2005.

[8] V. Krishnamurthy and D. Evans, "Hidden Markov Model Multiarm Bandits: A Methodology for Beam Scheduling in Multitarget Tracking", *IEEE Trans. on Signal Processing*, vol. 49, pp. 2893-2908, Dec. 2001.

[9] N. Visnevski, V. Krishnamurthy, S. Haykin, B. Currie, F. Dilkes, and P. Lavoie, "Multi-function radar emitter modelling: A stochastic discrete event system approach", *IEEE Conf. on Decision and Control*, pp. 6295-6300, Maui, Hawaii, USA, Dec. 2003.

[10] N. Visnevski, S. Haykin, V. Krishnamurthy, F. Dilkes and P. Lavoie, "Hidden Markov models for radar pulse train analysis in electronic warfare", *ICASSP-2005*, Philadelphia, March 2005.

[11] N.A. Visnevski, F. Dilkes, S. Haykin and V. Krishnamurthy, "Non-self-embedding context-free grammars for multi-function radar modeling--electronic warfare application", *IEEE International Radar Conference*, Washington, DC, June 2005.

[12] R.G. Wiley, *Electronic intelligence: the analysis of radar signals*, 2nd ed., Norwood, MA: Artech House, 1993.