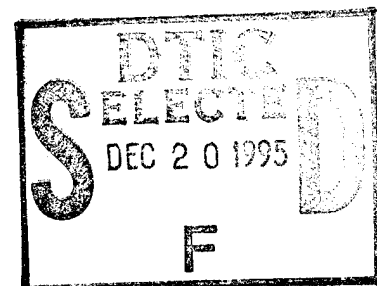# R & D Status Report

ARPA Order: 7684

Program Code: 1E20

Contractor: Brown University

Effective Date: June 1, 1991

Expiration Date: May 31, 1995

Contract Number: F30602-91-C-0041

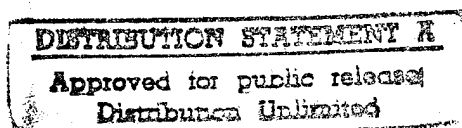Principal Investigator: Thomas Dean (401) 863-7645

Program Manager: Lou Hoebel (315) 330-7794

Title: Distributed Planning and Control for Applications in Transportation Scheduling

Date: June 1

Reporting Period: May 1, 1995–May 31, 1995

Sponsored by the Air Force and the
Advanced Research Projects Agency
under Contract No. F30602-91-C-0041

19951218 048

95-5-4d16

IN REPLY
REFER TO

DTIC-OCC

SUBJECT: Distribution Statements on Technical Documents

TO: ARPA/T10
ATTN: Ms. Cox
3701 North Fairfax Drive
Arlington, VA 22203-1714

1. Reference: DoD Directive 5230.24, Distribution Statements on Technical Documents, 18 Mar 87.

2. The Defense Technical Information Center received the enclosed report (referenced below) which is not marked in accordance with the above reference.

R&D Status Report
1 May 95 - 31 May 95
F30602-91-C-0041

3. We request the appropriate distribution statement be assigned and the report returned to DTIC within 5 working days.

4. Approved distribution statements are listed on the reverse of this letter. If you have any questions regarding these statements, call DTIC's Cataloging Branch, (703) 274-6837.

FOR THE ADMINISTRATOR:

1 Encl

GOPALAKRISHNAN NAIR
Chief, Cataloging Branch

FL-171
Jul 93

DISTRIBUTION STATEMENT A:

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

DISTRIBUTION STATEMENT B:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY;
(Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED
TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT C:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS;
(Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED
TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT D:

DISTRIBUTION AUTHORIZED TO DOD AND U.S. DOD CONTRACTORS ONLY; (Indicate Reason
and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT E:

DISTRIBUTION AUTHORIZED TO DOD COMPONENTS ONLY; (Indicate Reason and Date Below).
OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT F:

FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date
Below) or HIGHER DOD AUTHORITY.

DISTRIBUTION STATEMENT X:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS
OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE
WITH DOD DIRECTIVE 5230.25, WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC
DISCLOSURE, 6 Nov 1984 (Indicate date of determination). CONTROLLING DOD OFFICE IS (Indicate
Controlling DoD Office).

The cited documents has been reviewed by competent authority and the following distribution statement is
hereby authorized.

_A_
(Statement)

CASB / TIO
(Controlling DoD Office Name)

ARPA/TIO
3701 North Fairfax Drive
Arlington. VA 22203-1714

Unlimited
(Reason)

(Controlling DoD Office Address,
City, State, Zip)

Alba L. Clark
(Signature & Typed Name)

ARPA
(Assigning Office)

(Date Statement Assigned)

DEC 14 1995

MEMORANDUM FOR DTIC-OCC
                   ATTENTION: (GOPALAKRISHNAN NAIR)

SUBJECT: Distribution Statements on Technical Documents

This is in response to your memorandum requesting a distribution statement for the document entitled "Distributed Planning and Control for Applications in Transportation Scheduling." All copies should carry the Distribution Statement "A" (Approved for Public Release; Distribution is Unlimited).

Debra K. Amick
Technical Information Officer

Attachments

Status Report for May 1995

This is the last of the monthly reports for this contract (F30602-91-C-0041). Including the one-year extension, that makes this the 36th report. According to Lou Hoebel we are next in line for renewal if the freeze on ARPA funding is lifted. We were hoping to fund two graduate students and one post doctoral student for the summer. However, if we don't hear something soon, Shieu-Hong Lin, Lloyd Greenwald, and Jak Kirman will have to make other arrangements for the summer.

The research that we carried out for this contract has been quite successful, and efforts will continue to refine and distribute the software and algorithms that we developed during the past 36 months. Perhaps it is customary to spend the final report boasting of successes, but our results are chronicled in numerous conference papers, journal articles and book chapters. I think it might be more appropriate to give you some idea of the new directions that we are heading in as we approach this milestone in our research.

For the past nine months one of us (Tom Dean) has been visiting labs in the Western part of the country (University of Washington, Oregon State, University of Oregon and CIRL, University of British Columbia, Stanford, Berkeley, Rockwell (both the Palo Alto and Thousand Oaks science centers), University of Southern California and ISI, Los Alamos National Laboratories, and the Santa Fe Institute). One thing we have repeatedly learned is that artificial intelligence as a field is provincial in its technological breadth and relatively primitive in terms of its mathematical sophistication. During the period of this contract, we have studied and incorporated several new mathematical methods into our analysis and algorithm designs. These methods range from mathematical programming in operations research and nonlinear optimization techniques from mechanical engineering to time-series analysis in physics and related disciplines.

These studies have both increased our effectiveness (new tools and new tricks) and made it possible for us to communicate our results and our problems to a wider audience of researchers. At the Santa Fe Institute, we were able to quickly sift through the research concerning nonlinear dynamical systems to find techniques of use in our work on optimizing and learning dynamical systems for problems involving complex dynamics such as transportation scheduling. Our search spread through the Internet to reach researchers like Herbert Edelsbrunner (University of Illinois), Andrew Moore (Carnegie Mellon), Stefan Bornholdt (Universitaet Kiel), Bernard Chazelle (Princeton), Yoshua Bengio (Montreal), and many others from diverse backgrounds. As a consequence of our studies, we are putting together a proposal for a Spring Symposium on large dynamical systems with Melanie Mitchell (Santa Fe Institute) and Jim Crutchfield (Berkeley) and pursuing joint work with Moises Goldszmidt and Nir Friedman at Rockwell Palo Alto Research. The primary focus of this

new research effort is to develop better methods for learning dynamical system models. Ready access to such models has presented 'knowledge acquisition' bottleneck in our work on controlling dynamical systems corresponding to planning and scheduling problem.

In the last week, we have been investigating the use of techniques from statistical mechanics to solve very large optimization problems. There is an interesting phenomenon concerning the difficulty of optimization problems that as the number of relevant entities (aircraft, trucks) increases assumptions of uniformity can play an increasingly important role in reducing complexity. Even in medium-scale problems such as those faced by the military in transportation control and cities in highway traffic control, statistical techniques that rely on aggregation and (quasi) uniform behavior appear to be effective. We are particularly interested in hybrid methods that combine techniques from statistical mechanics and combinatorial optimization techniques that do differentiate with regard to local behavior.

In the remainder of this report, we relate some of our recent enquiries with regard to learning dynamical systems from data. Some discussion of this area was included in the previous month's report, but the following discussion pertains specifically to techniques that come from nonlinear dynamical systems theory. Hopefully, you will find it interesting.

Inferring Dynamical Systems: A Time Series Approach

We are interested in learning models of dynamical systems from data. In particular, trying to infer compact, stochastic finite automata consistent with the observed data. In work with Dana Angluin (Yale) and Jeff Vitter (Duke), we investigated the problem of recovering the structure of the underlying dynamical system (the state transition diagram) by actively 'exploring' the automaton. Exploring consists of moving about on the automaton by selecting inputs (the labels of edges) and observing the outputs of the resulting states. In the problems we are concerned with, both observations and state transitions are stochastic. These problems were motivated by navigation and map learning problems in robotics, and the results have a PAC flavor (the probably approximately correct model of Les Valiant).

In the robotics applications, states correspond to discrete locations and the size of the robot's environment is measured in terms of the graph of locations. In these applications, it is reasonable for the algorithms to be polynomial in the size of the state space. In our work for the planning initiative, we have been looking at much more complicated dynamical systems such as those governing transportation systems (air traffic, and urban package delivery). These applications have very large state spaces, but they often have lots of structure.

Suppose that there are N state variables, where a state variable might correspond to the location of a vehicle, whether a piece of material handling equipment is currently in

use, or whether a gate in an airport terminal is occupied by an aircraft. The state space is at least 2^N in the case of binary variables and most of our variables are not binary (though they are discrete). Let $f(y_t) = y_{t+1}$ represent a state transition function which is often considered as a random function, $\Pr(f(y_t)|y_t)$. In many practical problems, the state transition function f can be 'factored' into N simpler functions, one for each state variable, each of which is a function from M state variables, where M is a constant and $M \ll N$. We have $f(x) = \langle g_1(x), \ldots, g_N(x)\rangle$ such that each of the g_i can be represented using a table of size 2^M and the entire state transition function can be compactly represented in $O(N2^M)$ space.

Now suppose that we don't know this factored representation. Suppose that we only have a sequence of observations concerning flights into and out of a given airport. Can we infer a stochastic model that will allow us to make reasonable predictions concerning the arrival times for aircraft originating from other airports? There is also the related control problems (*e.g.*, gate or crew scheduling), but, in this report, we focus on simply learning the dynamics.

There are a number of methods we might use for learning the dynamics. Delay-coordinate embedding is the method of choice among many in time-series analysis. A closely related method for stochastic processes is the Baum-Welsh method for inferring hidden Markov models. Common to both of these methods is the idea of constructing a model whose state space consists of *lag* vectors corresponding to sequences of observations, $\{\langle x_t, \ldots, x_{t+r}\rangle | t = 1, \ldots\}$, where $x(y_t) = x_t$ is the observable output of the dynamical system at time t. (There is an appendix at the end of this report with a somewhat more detailed introduction to delay-coordinate embedding techniques. In particular, the appendix provides a clearer explanation of terms such as "attractor" and "diffeomorphism" which are used in the following discussion.) It can be shown, for example, that under a wide range of conditions the attractor (subspace) described by the dynamics of the actual system is diffeomorphic to the trajectory (subspace) described by the lag vectors corresponding to (suitably long) sequences of observations. The required length for lag vectors can be bounded by the dimension of attractor of the underlying dynamical system.

Now suppose that the the underlying dynamical system can be represented as a state transition diagram in which the states are simply labeled $1, \ldots, 2^N$ (again assuming boolean variables for simplicity). Is the N-dimensional structure of the state space apparent in any properties of the graph corresponding to the state-transition diagram? What is a reasonable measure of the 'dimension' of the underlying attractor in this discrete version of the problem? Is there a useful analog of the correlation dimension[1] for discrete stochastic dynamical systems?

---

[1] The correlation dimension is obtained by analyzing the correlation of a set of points that has been moving on an attractor for some time. The correlation dimension weights the points on the attractor according to

We are searching for useful methods of categorizing such dynamical systems. These methods might be useful in two ways. First, we may be able to devise efficient approximation algorithms (polynomial in N) for particular classes of problems. Second, we may be able to estimate the 'dimension' of an attractor from data and use this information to expedite inference (say, by determining the length of an appropriate lag vector).

There are technical problems in integrating ideas from nonlinear dynamical systems where dynamics are represented as differential maps and ideas from computer science where dynamics is discrete time and space, but perhaps not as difficult problems as your might imagine. Oddly enough, nonlinear dynamics is typically interested in noisy observations of complex (read possibly chaotic) deterministic systems. The interplay between deterministic and stochastic systems (in which the aim is to recover the relevant distributions) is subtle. Physicists assume that the observations are discrete and bounded both in precision and range; this implies that our understanding of physical systems is necessarily in terms of a discrete approximation of the real world. One of the most difficult challenges concerns translating the language of smooth mappings and continuous manifolds into the language of graphs.

The notion of dimension mentioned earlier is a prime example of the difficult of translating ideas from differential geometry to discrete mathematics. Differential geometry is the source of many of our intuitions about real physical phenomena; it is critical that we be able to apply those intuitions in the case of discrete models. The transitions in a state transition graph are simply edges and therefore 1-dimensional. In order to apply ideas from differential geometry we have to introduce higher dimensional objects involving 3 and more states at a time. These objects would then form triangles or higher-dimensional simplices. In addition, we would have to introduce some notion of distance in order to get to a topological or geometric interpretation. It is currently not clear how this approach will work out mathematically, but it is definitely interesting as a means of introducing powerful techniques to solving combinatorial problems. As usual, pointers and suggestions are welcome.

Recent Publications

Adnan Darwiche and Tom Dean are in the process of completing work on a video describing recent advances in probabilistic diagnostic systems. We hope to have a copy of this video for the Planning Initiative Workshop meeting in San Diego in June. This video was made with help from Rockwell International and sponsored by the American Association for Artificial Intelligence.

---

how frequently they are visited (in contrast to the geometric dimension of the attractor in which all points are weighted identically). For a continuous attractor, if the correlation dimension is much greater than 1 and is not an integer, then this is an indication that the attractor is strange and the dynamics are chaotic.

Postscript versions of our recently accepted papers in IJCAI and UAI are available on line using the following URLs.

ftp://ftp.cs.brown.edu/u/tld/postscript/DeanandLinIJCAI-95.ps
ftp://ftp.cs.brown.edu/u/tld/postscript/LittmanDeanandKaelblingUAI-95.ps

You also might be interested in an article on planning and scheduling which will be published in the CRC "Handbook on Computer Science and Engineering." This paper was written by Tom Dean and Subbarao Kambhampati (Arizona State). Postscript for this article can be found using the following URL.

ftp://ftp.cs.brown.edu/u/tld/postscript/DeanandKambhampatiCRC-95.ps

Appendix: Delay-Coordinate Embedding

In this appendix, I'll try to summarize what I know about delay-coordinate embedding techniques as they are used in time-series prediction and nonlinear dynamical systems theory. I was first introduced to these techniques by Paul Dagum at Rockwell Palo Alto Research who used the delay-coordinate embedding to monitor and predict respiratory crises for infants with ARDS (adult respiratory distress syndrome). (There is no agreed upon physiological model for ARDS. To complicate matters, the time-series analysis had to be done with data gathered while patients were undergoing treatment. Much as the data would be gathered say in learning the dynamics for a transportation problem. The task was ultimately to generate a better treatment policy. There was no attempt to construct a physiological model; the goal was simply to predict the evolution of physiological measurements as a consequence of specific treatments.) When I got to the Santa Fe Institute, I found that lots of people were using these techniques, from the people doing time-series prediction (Farmer, Lapedes, Casdagli) to the emergent computation people (Crutchfield, Mitchell, Hanson).

The basic idea is quite elegant and almost too good to be true. As with many such ideas, its application in practice is quite complicated. Nevertheless, delay-coordinate embedding is the method of choice in many time-series prediction applications. The idea is as follows. Assume we have a sequence of measurements of some unknown dynamical system whose governing equations are specified by $dy/dt = f(y)$, $y \in R^n$, and an output function $x(y)$. In particular, we have a series of time-delayed observations, $\{x_t = x(y_t)\}_{t=0}^{\infty}$. The delay between measurements is fixed and denoted by τ. These criteria are met for example in the observations of many transportation problems.

Takens [1980] showed that for some integer d the space of (lag) vectors $x = \langle x_t, \ldots, x_{t+d} \rangle$ is *diffeomorphic* (i.e., the behavior of x and y differ by a smooth local invertible change

5

of coordinates, referred to as an embedding) to the solution space of f, for almost every possible choice of f, x, and τ as long as d is large enough, x depends on at least some components of y, and the remaining components of y are coupled by the governing equations to the ones that influence x.

Originally, Takens applied his results to detecting (inferring) attractors in turbulence (a classic example of a chaotic dynamical system). Physicists use the term 'attractor' to refer to the asymptotic, steady-state behavior of a physical system, $e.g.$, the behavior of a stochastic system once it has entered an ergodic subset of the state space. In particular, the initial transient behavior of the system is ignored. This attractor may be complicated, involving several 'strange' attractors (transient phases corresponding to quasi-stable cycles that the system switches between in a manner that is often difficult to predict).

Obviously, there are degenerate choices for τ for which the embedding fails. But these choices are of measure zero in most cases and reasonable methods of sampling and intrinsic noise in observation avoid problems in practice. There is a large body of theory and practice regarding how to choose d and its connection to the dimensionality of the underlying manifold (the solution space of f). Once you have chosen an appropriate delay τ and embedding dimension d, tracing out the attractor and extrapolating from the resulting map is relatively straightforward. Most practical delay-coordinate embedding methods involve preprocessing the data using some pretty sophisticated filters, but this preprocessing is orthogonal to the issues concerning the application of delay-coordinate embedding.

Techniques similar to delay-coordinate embedding have been used to infer finite state models of dynamical systems. Crutchfield and Young's [1989] work on ϵ-machines is one such example. Crutchfield and Young talk about the general task of machine reconstruction. An ϵ-machine is a machine reconstructed from a time-series of observations whose precision is related to ϵ. The embedding for an ϵ-machine is defined on a d-dimensional discrete lattice of cells of size ϵ. Inference proceeds by fixing d (the width of a window on the data) and building a tree to represent states of the dynamical system. The methods are similar to those used in constructing a code book in data compression. If d is chosen appropriately the resulting code book allows you to assign names to states of the machine and reconstruct the transition table. These methods can be extended to reconstruct the full probability distribution for a stochastic process.

The method of machine reconstruction described above will fail for cases in which the data is not generated by a finite state process. Crutchfield and Young describe results concerning the reconstruction of machines modeling dynamical systems based on the logistic function at the onset of chaos. In this case, there is a phase transition such that the dynamics changes from that which can be modeled by a finite state machine to that requiring a stack-based machine ($e.g.$, a push down automaton). Such phase transitions are

6

of particular interest in what has become called "emergent computation." The method of Crutchfield and Young is remarkably similar to the idea of using distinguishing sequences to identify states in learning automata which we employed in [Dean *et al.*, 1995]. We hope to use some of our insights regarding the computational complexity of learning such distinguishing sequences to understand more about the methods of Crutchfield and Young.

A natural question to ask is whether methods such as machine reconstruction provide any insight into learning stochastic models (recovering the distribution governing state transitions) in the case of discrete-time and discrete-space processes. It may be that the essential insights are already a part of the literature on compression and hidden Markov models. However, there are a variety of specialized techniques that seem to be of immediate relevance to the problem of inferring probabilistic models based on Bayes networks [Heckerman, 1994].