NASA
Contractor Report 195380

Army Research Laboratory
Contractor Report ARL–CR–187

/4 P.

Intelligent Control Systems Research

Kenneth A. Loparo
Case Western Reserve University
Cleveland, Ohio

September 1994

Prepared for
Lewis Research Center
Under Grant NAG3–788

National Aeronautics and
Space Administration

U.S. ARMY

RESEARCH LABORATORY

Research on Intelligent Control, supported by the NASA Lewis Research Center and the U.S. Army, has been conducted by the Department of Systems Engineering at Case Western Reserve University. This work began in 1987 with an initial research contract to support a literature survey and problem formulation for the concept of intelligent control. Several questions were asked in the earlier work in an attempt to focus on concepts and ideas that would be relevant for future research studies. During the initial period of this work a detailed report on the methods and techniques from systems and control theory, hierarchical and multilevel systems theory, expert.systems and AI, learning systems and automata theory, which were relevant to the general area of intelligent control was prepared. Also, a study was conducted to determine the performance of humans in control tasks.

The experiment was based on a computer simulation of the classical pole balancing problem, where a concentrated mass is located at the end of an inverted rod attached to a cart which can move on a horizontal surface. The simulation included various methods of representing the system data to the operator. For example, in one simulation the pole and cart system was graphically displayed and the operator could view the time evolution of the system as forces were applied to the cart. In another operating mode, a bouncing ball whose frequency was proportional to the velocity of the pendulum was displayed. The operator had to discover by trial and error which direction of the ball's motion was associated with clockwise and counterclockwise motion of the pendulum; a failure condition was always given to the operator when the pendulum would fall to the horizontal position. For the experiment, disks which contained this experiment were distributed to different people, some technical and some non technical, to determine the ability of these individuals to "learn" an appropriate control law. The level of force and the initial angular perturbation of the pendulum were randomized over the simulation runs. All the control moves of the participants were recorded and then analyzed at a later date.

The conclusions from the experiment were not surprising and formed the basis for the technique of "learning" or "intelligent" control that we adopted for the first year of our research work: a *reinforcement learning approach.* In a reinforcement learning

approach to control, there is a discrete set of control alternatives and a performance functional which is used to evaluate the effectiveness of the control inputs. The state space of the dynamical system that is to be controlled is quantized into a collection of sets called *situations,* and the objective of the reinforcement learning controller is to assign to each quantized set a "unique" control value which is preferred on the basis of the performance functional. The reinforcement learning method uses a reward and penalty scheme to adjust a set of probabilities, with one probability associated with each control input/situation pair.

Implementation of the reinforcement learning controller is at the *direct control level* of the intelligent control hierarchy. In the problem formulation developed in our first year effort, we proposed an intelligent control hierarchy which utilized a functional decomposition of the overall control problem. This decomposition included: a *direct control level* that was responsible for responding in real time to disturbances in the plant, a *planning/optimizing level* controller which is responsible for modifying set points, parameters and performance goals for the direct control level to respond to changes in the plant or operating environment and a *supervisory/explanation facility* and *user interface* to the intelligent control system which provides the operator with qualitative information and explanations about the process and the performance of the intelligent control system. The operator (control system user) can use the information supplied by the explanation facility to modify process knowledge and goals. This functional decomposition is commensurate with a temporal decomposition of the control tasks-as the complexity of the decision/control problem increases, along with the computational time required to determine the appropriate control action or decision, the control task is relegated to higher levels of the intelligent control hierarchy.

The major effort for the first phase of the research work was in the implementation and evaluation of the direct level controller. The direct level controller incorporates six subsystems for learning/control selection. The *critic* is the evaluation subsystem in the direct level controller. This subsystem accepts output data from the process and the control database and provides a reinforcement (reward)/punishment signal to the *learning* subsystem. An important problem that was addressed in the design of the critic was the *credit assignment problem.* As we are dealing with a dynamical system, there is a functional relationship between

2

the process inputs and outputs. Hence, the critic must know how to assess credit or blame to past and current controls based on the current value of the process output. We developed specific techniques to deal with this complexity and the details can be found in [1] or [2]. The *learning* system uses reinforcement data from the critic to adjust the (conditional) probabilities of the situation/control pairs; a linear-reward-penalty scheme is used in the implementation. The learning system computes an update of the situation/control probabilities and provides this information to the *control database* subsystem. Before a control for the current time period can be computed, the process output must be analyzed to determine the "state" of the system. This involves three subsystems of the direct control level, the *data monitor,* the *situation recognition unit* and the *control selection unit.* The data monitor is analyzing the process output to determine anomalous conditions, such as sensor malfunctions, which will affect the quality of the data and the performance of the direct level controller. If the data monitor passes the output data, it is classified into situations in the situation recognition unit. The output space of the process is quantized into sets referred to as situations, and the situation recognition unit assigns a situation number to the observed process output.

Remark: Quantizing the process output into situations can be difficult and can induce complicated behavior in the controlled system. The problem is that the direct level controller is attempting to assign a unique control value to each situation. However, the dynamics of the quantized system can be quite complicated and, in fact, in some instances the evolution of the quantized system is random [3 and 4]. In such cases, the learning unit is unstable in the sense that controls which are rewarded for a particular situation at one time are penalized for the same situation at another time. This is a direct result of the fact that the output quantization for the process does not necessarily define a Markov partition for the system's output flow.

The direct level controller is operating in a closed-loop configuration. In such cases, it is well known that identification (learning) and control can compete; this is referred to as the *dual control effect.* The problem stems from the fact that if the controller is doing a good job regulating the plant, then presumably the output of the process remains in a neighborhood of the desired set point or trajectory, and the input/output data which is collected

is not very informative about the general characteristics of the process dynamics-identification is difficult. This so-called dual effect is also a problem in the direct controller where learning and control are occurring simultaneously.

The direct level controller was implemented in a Texas Instruments Explorer System and tested in simulation for the control of an inverted pendulum. This particular problem was chosen because it has been used in past experimental (simulation) studies to evaluate different methods of intelligent or learning control. Although the direct level controller showed reasonable performance in learning a stabilizing controller for the inverted pendulum in a variety of different operating configurations, on many occasions the learning times required by the controller were prohibitively large and the control probabilities would exhibit oscillatory behavior. A detailed analysis of the phenomena led to the conclusion that it was the quantization of the output space into situations and the dual effect of the combined learning/controller synthesis that were the root causes. These problems were addressed in detail in the second year research work.

The second phase of the research effort concentrated on developing a refined implementation of the direct level controller, including an adaptive/optimizing level function for the learning phases of the controller. As mentioned previously, the direct level controller uses a reinforcement learning control paradigm to synthesize the control action. The control actions are rewarded if they improve the dynamical behavior of the system as measured by a performance functional termed the *subgoal*, and punished otherwise. The problem of determining an appropriate subgoal for the instantaneous evaluation of the performance of the system, derived from the overall performance functional for the process which is being controlled is system dependent and, in general, is unsolved. In this work we have used a heuristic approach to construct a subgoal for the problem of stabilizing the inverted pendulum. No general results for arbitrary systems have been determined.

The direct level controller operates as developed in the phase one research effort. The adaptive/optimizing level is developed to improve the operation of the direct level controller by adjusting the information classifying scheme. The reinforcement learning control scheme decomposes the control action synthesis task into: (1) classifying the input/output data of the process into situations and,

(2) determining the control action which maximizes the a posterior probability of being the correct control action for the situation identified. The controller has two objectives; learn as much as possible about the plant and synthesize the best control policy as measured by the performance functional for the plant. As in a classical adaptive control scheme, the learning and control objectives are usually competing. These two objectives are used to distinguish between two distinct phases of the learning process.

During learning, classification of measured data from the plant into situations is based on neighborhoods defined in the input/output space of the process. The neighborhoods are induced by a similarity metric and the learning process is decomposed into two phases: the *creation phase* and the *refinement phase*. In the creation phase, controls are applied randomly to the process in an attempt to stimulate all modes of the system and enhance identification/learning at the possible expense of control performance. In the refinement phase of the learning process, control actions are determined by their expected success in terms of a subgoal objective and the topology of the neighborhoods are altered in an attempt to find a partition of the output space of the process such that a unique control is associated with each input/output situation pair.

A unique feature of the work is the introduction of the concept of entropy as a means of guiding and evaluating the determination of the neighborhoods during the creation phase and the refinement phase. The creation phase is identified by the entropy (or uncertainty) of each neighborhood being greater than a given, user-specified, threshold. Once the entropy has been reduced to less than the threshold, the learning switches from the creation phase to the refinement phase. For more details the reader is referred to [5].

The adaptive/optimization level intervenes during phase two of the learning process based on observed *anomalies* in the direct level controller. The anomalies are either events which cause a particular control action to increase the entropy associated with a particular situation or a partitioning of the output space. The underlying concept of intervention is that by altering the topology of the neighborhoods, the partition of the output space of the process, the behavior of the learning control scheme can be improved. Although it is not always true, smaller neighborhoods usually improve the controller performance at the cost of additional computational

complexity. One possible intervention strategy is to adapt the threshold of the similarity metric which defines the neighborhoods. Adjustments of the threshold treats all directions in the output space uniformly and such a scheme can result in deterioration of the overall performance of the controller. Therefore we have chosen to use a parametric adjustment of the weighting matrix in the quadratic similarity metric which is used to classify input/output patters into situations. A gradient based algorithm is derived to provide adjustments to the similarity metric. Refer to [5] for details.

The final accomplishments of the work in phase two were refinements and enhancements to the implementation of the intelligent controller on the TI Explorer computer system. The windows environment and graphics capability of the Explorer system were exploited to develop a user interface. With this user interface and the incorporation of animation into the system makes it a suitable platform for development work in intelligent control.

The third and final phase of the intelligent control system research effort was aimed at relaxing some of the restrictions of the reinforcement/learning control paradigm which formed the basis of the direct level controller. Two approaches were taken during this work, both incorporating the use of a priori information into the realization of the intelligent control system. The first approach was to consider an alternative learning control method based on feedforward neural networks for a special class of nonlinear dynamical systems; the class of linear-analytic systems. The other approach was to use a priori system information to develop methods that would extend the capabilities of the reinforcement/learning control approach. We mentioned earlier the problem which results because of quantization of the output space of the process, the other problem is the quantization of the input or control space. This issue was studied as part of the third phase of the research work.

Linear-analytic systems are a general class of nonlinear systems where the control input enters linearly into the system dynamics and the vector field which defines the system flow when the input is fixed is made up of analytic functions of the state of the system. This class of systems is important for at least two reasons: (1) many nonlinear systems can be represented by, or approximated by, dynamical systems of this form, and (2) from this class of systems it is possible to develop a theory of control system

6

synthesis which closely resembles the well known linear theory. Our approach was to utilize the fact that for systems of the linear-analytic type, there exists a theory of control synthesis in which a feedback control is derived which linearizes the input/output dynamical behavior of the system. If we could develop an intelligent control structure that would learn the linearizing feedback controller, then classical linear control methods could be used on the linearized system to obtain the desired closed-loop system performance. The realization of the intelligent controller chosen for this part of the work was in terms of a feedforward neural network, where unsupervised learning methods were developed for this application to guide the selection of an appropriate linearizing feedback control input.

We began with the assumption that the linear-analytic system was feedback linearizable and then used this information to select the appropriate form of a linear system which was used during training. This idea is similar to a model-reference adaptive control scheme, except in our implementation a feedforward neural network was used as the controller and a gradient based algorithm (an extension of the familiar back propagation algorithm for a feedforward neural network) was used to adjust the network parameters using real-time input/output data from the system. For more details on the theory and applications of this work the reader is referred to [7] and [8].

The alternative approach we investigated for incorporating a priori system information into the synthesis of learning control strategies was to focus attention on two dimensional systems and their geometric properties. As we mentioned earlier, one problem with reinforcement/learning schemes is related to partitions of the output or state space of the process to be controlled. For control problems related to set-point regulation, including stabilization, the existence of a suitable control which transfers an initial point to the desired final point is determined by the *attainability* and *reachability* properties of the system. Therefore, the ability of the learning control system to determine a suitable control action for a particular point-to-point steering control problem also depends on these geometric properties of the system. In this work we have used methods of characterizing the attainable and reachable sets for a dynamical system to enhance the performance of a learning control system. The attainable and reachable sets are parameterized by the control input which is assumed to be held constant over a fixed time

7

interval, referred to as the control time. This is consistent with a digital (discrete time) implementation of the controller where, for example, a zero order hold would be used as a reconstruction device. For a single input system, we assume the control takes values in a compact convex subset (an interval) of the set of real numbers (R) and the control set includes the origin. Given an initial point p, we define the attainable set from p (*A(p)*) on the interval [0,T] to be the collection of trajectories of the controlled system initial from p, given that the control input ranges over the set of admissible control inputs. Similarly, given a target (final) point p, we define the reachable set (*R(p)*) on the interval [0,T] to be the collection of trajectories of the controlled system which can be steered to p as the control input ranges over the set of admissible control inputs. The attainable and reachable sets play important roles in problems related to point-to-point steering in control systems. The geometry of the sets *A(p)* and *R(p)* depends on the characteristics of the system and the set of admissible controls. For more details refer to the thesis [9] and the papers [10] and [11].

The problem of intelligent control as formulated in this work is to learn an appropriate feedback control strategy which will steer a given set of initial points to a given final point on a time interval [0,T]. Of the difficulties we encountered with reinforcement/learning control in our previous years' work, the discretization of the control set and its influence on the dynamical system performance was a focus of this research effort in the final year of the project. An importsnt issue is that in order to have "fine" control of the system the number of partitions of the control set (i.e. the number of control values) must be large, but this causes computational and numerical problems in the reinforcement/learning algorithms. Using a priori information about the system-the geometry of the attainable, reachable and admissible control sets-we developed an adaptive form of the reinforcement/learning control suitable for a broad class of nonlinear two dimensional systems. The basic theory behind the method is to use the convexity property of controllable sets S in the phase space of a two dimensional nonlinear system. In this set S, all points are attainable and reachable with respect to all other points in the set and the boundary of the set S is determined by extremal trajectories of the controlled system. That is, for the case of a single input system, if the control set is the interval [a,b], then the extremal trajectories are determined by choosing the control to be equal to a or b, respectively. In planning a trajectory from an initial point p to a

target point t, we select a path in the phase space which consists of a collection of attainable and reachable sets which have pairwise nonempty intersections. If there is no such path, then the point-to-point steering problem has no solution. Then, a collection of extremal trajectories forms a boundary for this region and the learning algorithm attempts to synthesize an appropriate control sequence which will accomplish the desired transfer. Using convexity properties of the controllable sets, the algorithm iterates on the partition of the control set to continuously refine the partition while keeping the number of elements in the partition constant. In this way, we have developed an adaptive reinforcement/learning scheme which has essentially a continuum of control values. There is a course partition of the control set which includes the extremal controls, and at each iteration based on input/output data from the system and the geometric properties of the reachable and attainable sets, the control set partition is refined and the learning is continued. More details and a simulation study can be found in the thesis [9].

This research program has been very productive and a number of important issues in intelligent control have been identified and a number of important contributions to the theory and application of intelligent control methods have been made. Significant contributions include:

1. The development of a hierarchical framework for intelligent control [1] and [2];

2. The development, implementation and testing of a direct level controller based on a reinforcement/learning control paradigm [1] and [2];

3. The development of an information-theoretic framework for adaptive learning to address the difficult "dual" effects of the reinforcement/learning controller [5] and [6];

4. The implementation of a adaptive/optimizing control layer within the intelligent control hierarchy for improving learning and control performance [5] and [6];

5. The development and implementation of a software based simulation and graphically based evaluation tool for use as an intelligent control system development tool [1] and [5];

6. The study of quantization effects on the dynamic system trajectories, including entropy measures and active probing, chaos and complicated dynamical system behavior [3],[4] and [5] and [6];

7. The application of neural networks for direct level control, the synthesis of feedback linearizing direct level controllers for linear-analytic systems using learning control methods [7] and [8];

8. The use of a priori system information in learning control system synthesis, reachable and attainable sets and an adaptive scheme for refining control set partitions to improve closed loop control system performance [9].

REFERENCES

[1] Berger, K., " A Hierarchical and Intelligent Controller for Complex and Uncertain Processes," M.S. Thesis, Department of System Engineering, Case Western Reserve University, Cleveland, Ohio 1988.

[2] Berger, K. and Loparo, K.A.," A Hierarchical Framework for Learning Control," in *Artificial Intelligence, Simulation and Modeling,* ed. L.E. Widman, K. A. Loparo and N.R. Nielson, Wiley, 1989.

[3] Feng, X. and Loparo, K.A., " Active Probing for Information in COntrol Systems with Quantized State Measurements: A Minimum Entropy Approach," to appear *IEEE Transactions on Automatic Control.*

[4] Feng, X. and Loparo, K.A., A Study of Chaotic Behavior in Linear Discrete Time Sytems with Quantized Measrements, submitted to *IEEE Transactions on Automatic Control,* also in *Proceedings of the 31st IEEE Conference on Decision and Control,* Tucson, Arizona, December 1992, pp. 2107-2112.

[5] Musgrave, J.L., " An Information Theoretic Approach to State Classification in Reinforcement Learning Control," M.S. Thesis, Department of Systems Engineering, Case Western Reserve University, Cleveland, Ohio, 1989.

[6] Musgrave, J.L. and Loparo, K.A., " Entropy and Outcome Classification in Reinforcement Learning Control," *Proceedings of the IEEE Conference on Intelligent Control,* October 1989.

[7] Hassibi, K., " A Study of the Application of Neural Networks to Feedback Linearization," Ph.D. Dissertation, Department of Electrical Engineering and Applied Physics, Case Western Reserve University, Cleveland, Ohio, 1991.

[8] Hassibi, K. and Loparo, K.A., " Application of Neural Networks to Feedback Linearization," to appear *IEEE Transactions on Sytems, Man and Cybernetics.*

[9] Prabhu, K.D., " Learening Control for Two Dimensional Nonlinear Systems," M.S. Thesis, Department of Systems Engineering, Case Western Reserve University, Cleveland, Ohio, 1992.

[10] Fashoro, M., Hajek, O. and Loparo, K.A., " Reachability Properties of Constrained Linear Systems," *Journal of Optimization Theory and Applications,* Vol. 73, No. 1, April 1992, pp.169-195.

[11] Fashoro, M., Hajek, O. and Loparo, K.A., " Controllability Properties of Constrained Linear Systems," *Journal of Optimization Theory and Applications,* Vol. 73, No. 2, May 1992, pp.329-346.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1994	3. REPORT TYPE AND DATES COVERED Final Contractor Report

4. TITLE AND SUBTITLE

Intelligent Control Systems Research

6. AUTHOR(S)

Kenneth A. Loparo

5. FUNDING NUMBERS

WU–505–62–0E
1L161102AH45

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Case Western Reserve University
Department of Systems Engineering
Cleveland, Ohio 44106

8. PERFORMING ORGANIZATION REPORT NUMBER

E–9113

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Vehicle Propulsion Directorate
U.S. Army Research Laboratory
Cleveland, Ohio 44135–3191
and
NASA Lewis Research Center
Cleveland, Ohio 44135–3191

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA CR–195380
ARL–CR–187

11. SUPPLEMENTARY NOTES

Project Manager, Jonathan S. Litt, Vehicle Propulsion Directorate, NASA Lewis Research Center, organization code 0300, (216) 433–3748.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 63

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Results of a three phase research program into intelligent control systems are presented. The first phase looked at implementing the lowest or direct level of a hierarchical control scheme using a reinforcement learning approach assuming no a priori information about the system under control. The second phase involved the design of an adaptive/ optimizing level of the hierarchy and its interaction with the direct control level. The third and final phase of the research was aimed at combining the results of the previous phases with some a priori information about the controlled system.

14. SUBJECT TERMS

Intelligent control; Hierarchial control; Artificial intelligence

15. NUMBER OF PAGES 13

16. PRICE CODE A03

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102