Performance Metrics for Intelligent Systems

John M. Evans and Elena R. Messina
Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899-8230

ABSTRACT

Research into intelligent systems and intelligent control is burgeoning. However, there is no consensus on how to define or measure an intelligent system. This lack of rigor hinders the ability to measure progress in the field and to compare different systems' capabilities. We discuss some of the challenges and issues in defining performance metrics for intelligent systems and issue a call to action to participants in the Performance Metrics for Intelligent Systems Workshop to define practical metrics that will advance the state of the art and practice.

KEYWORDS: *performance metrics, intelligent systems, intelligent control*

1. INTRODUCTION

Intelligent systems are increasingly being identified as solutions to many advanced applications in manufacturing, defense, and other domains. Industry workshops [4] and roadmaps [3] specifically call for intelligent control or intelligent systems to address needs such as

- Adaptive, reconfigurable manufacturing equipment and processes
- Self-optimizing, science-based control of manufacturing unit processes
- "First part correct," that is, the ability to design and manufacture a product correctly, the first time and every time
- Self-diagnosing and self-maintaining systems
- Tool wear and breakage monitoring

Government agencies are basing major programs on intelligent capabilities, for example,

- The Army Experimental Unmanned Ground Vehicle Systems (Demo III)
- Defense Advanced Research Projects Agency (DARPA)/Army Future Combat Systems
- DARPA Mobile Autonomous Robot Software
- DARPA Software for Distributed Robotics
- DARPA Tactical Mobile Robots
- National Aeronautics and Space Administration (NASA) spacecraft and rovers
- Department of Energy (DOE) waste remediation robot systems
- Department of Transportation (DOT) Intelligent Vehicle Initiative

In addition to the examples above, there are myriad other efforts in academia, industry, and government labs of work referred to as "intelligent systems." Despite the common use of "intelligent system" and "intelligent control," there is no uniform definition for either term. Generally, they are characterized by having one or more of the following traits [1]:

- Adaptive
- Capable of learning
- "Does the right thing" or "acts appropriately"
- Non-linear
- Autonomous symbol interpretation
- Goal-oriented
- Knowledge-based

These terms are ambiguous and qualitative. The Intelligent Systems Division of the National Institute of Standards and Technology has launched an initiative to better define what an intelligent system is and how to measure its performance. The mission of the Intelligent Systems Division, one of five divisions in the Manufacturing Engineering Laboratory, is "to develop the measurements and standards infrastructure needed for the application of intelligent systems by manufacturing industries and government agencies."

We are working with various industry groups and government agencies to tackle the issue of intelligent system performance. The Performance Metrics for Intelligent Systems Workshop is a foundational step, which brings together a multi-disciplinary community to help define the highest priority areas to concentrate on, having the highest payoff.

2. THE CHALLENGE OF DEFINING AND MEASURING MACHINE INTELLIGENCE

Researchers have been pursuing forms of machine intelligence for several decades. There have been many areas of focus, such as natural language understanding, expert systems to aid diagnoses, and decision-making tools for financial systems. Closer to our domain of interest, much effort has been focused on defining intelligent control as a discipline, but even so, there are no quantitative measures.

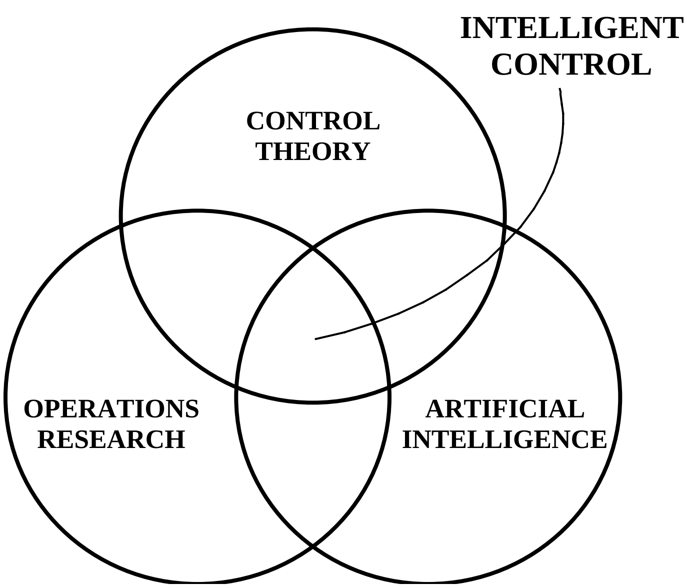

Figure 1: Intelligent Control as of 1985

Beginning with the efforts of Fu [1] and Saridis [3] in the seventies, there have been numerous conferences and workshops aimed at the topic of intelligent control. Nevertheless, the field remains fragmented due to its multidisciplinary nature. As noted in the first Symposium on Intelligent Control in 1985, intelligent control was proclaimed a theoretical domain, in which control theory, AI, and operations research intersected (Fig. 1 from [6]).

The definition of an intelligent system may be considered broader than that of intelligent control. As a "system," there may be more constituent parts, such as perception, world modeling, or value judgement. Yet more disciplines are brought into the picture. Examples of these include data representation, image processing, and decision theory.

Given the multi-disciplinary nature of the systems we are concerned with, it is clear that *defining the scope and performance of these systems is a challenge.* Terminology is one of the first hurdles that must be overcome. Different disciplines ascribe different definitions to the same words. For example, "complexity" may refer to non-linear systems in one field and to computational resources needed in another.

It is very difficult, if not impossible to currently evaluate research into intelligent systems. Since there are no quantitative metrics, intercomparisons of results are not generally possible. Sponsors are not able to adequately judge whether research results meet their requirements. Potential users have no impartial evaluation reports, *a la* "Consumer Reports," of intelligent systems, techniques, and tools. In general, the lack of metrics slows progress. There is a proliferation of data specific algorithms and task-specific solutions.

One of the biggest costs paid is the duplication of effort. New programs may be unable to have a firm definition of past accomplishments, hence they may fund work that repeats previous research. Research teams cannot leverage prior existing work from other institutions and tend to have to reinvent the wheel by building all of their system's components from scratch. They are burdened with having to spend effort in building components that are not part of their research focus, instead of being able to leverage existing "best of class" solutions and focussing on their interests.

Another negative impact, from the sponsor's viewpoint, is the lack of predictive ability in assessing new applications. Without objective performance evaluation metrics and an understanding of capabilities and limitations, it is difficult or impossible to assess claims of competing approaches in formulating new projects and programs. This leads to inefficiencies and failures that could be avoided if we had the measurement tools that we need.

3. Issues in Measuring Performance

Numerous questions must be answered when considering how to define the performance of these intelligent systems. We will present a few questions. Many more will arise as we delve into the matter more closely.

- Should we measure only the external behavior of a system? Is that the only aspect that can feasibly be measured? Or, is there value in decomposing a system into components and measuring their individual capabilities? Examples would be measuring the path planning algorithms in isolation from the perception and other control subsystems.

- How generic does the measure of a system's intelligence have to be? Should we strive for general intelligence metrics that are domain-independent or are we better off focussing on application and domain-specific metrics? Are domain-independent metrics even meaningful?

- How do we factor in "body intelligence," the mechanical capabilities of a system as opposed to the control capabilities, when assessing the performance of a system? If we have a mobile robot, some of its abilities to achieve its stated goal (e.g., traverse a rubble pile to find survivors) can be attributed to its mechanical properties rather than its software intelligence.

- Are testbeds a viable measure of performance, or do they invite "gaming," that is, encourage solutions that are tailored to performing well in the testbed? If we don't have testbeds, how can we achieve reproducible measures of performance?

4. Initial Observations

One of the complicating factors in discussing intelligent systems is the use of the word "intelligence." It is freighted with significance and analogies to human or biological intelligence naturally arise. The quest for standard, uniform measures of intelligence in biological systems remains a subject of controversy. Therefore, we would advocate avoiding the temptation to spend too much time striving for performance measures that are based on human or higher level biological systems.

Observing that we are dealing with multi-disciplinary technologies and multiple application domains, we should expect that no single, unique measure of performance is feasible. Therefore, no single overarching and generic intelligence test will suffice. We need to strive for the right granularity of metrics.

We must be prepared to attack the problem on multiple fronts. It probably won't suffice to have just a theoretical investigation or an experimental one. Research must proceed on the theory as well as on gathering experimental data.

One of the key attributes of intelligent systems is its multi-disciplinarity. This poses a challenge, but also an opportunity. We can come together from a variety of disciplines and form a new community in which we share our expertise. We must have dialog and information exchange amongst ourselves in order to synthesize the best results from the different fields that contribute towards intelligent systems research.

That is the purpose of this workshop and the reason for the diversity of the presentations that you will hear.

5. Call to action

The challenge is thus to define performance measures for new and evolving intelligent systems technologies that can greatly improve industrial productivity and advance government mission objectives. We must work together to build a technical foundation for measuring performance. This includes agreeing on the domains to investigate and a common set of terminology. We must develop theoretical foundations, methodologies, and supporting infrastructure for achieving our goals. Ultimately, measures must be developed that are practical, unambiguous, easy to use and widely deployable. We must simultaneously focus on attainable goals and strategies for both near-term and long-term measures of performance, as our understanding of them and the capabilities of the systems themselves evolve. Researchers, industry, and government will benefit from practical solutions they can readily apply, not from philosophical ones.

6. References

[1] P. J. Antsaklis, "Defining Intelligent Control", Report of the Task Force on Intelligent Control, P.J Antsaklis, Chair, IEEE Control Systems Magazine, pp. 4-5 & 58-66, June 1994.

[2] K.-S. Fu, "Learning Control Systems and Intelligent Control Systems: An Intersection of Artificial Intelligence and Automatic Control," IEEE Trans. on Automatic Control, Vol. AC-16, 1971.

[3] Integrated Manufacturing Technology Roadmapping Initiative, "”Manufacturing Processes and Equipment Draft Roadmap,” Oak Ridge, TN, December, 1998.

[4] Neal, R. and Messina, E., ed., Proceedings of the First Part Correct Workshop, Gaithersburg, MD, April 2000.

[5] G. Saridis, Self-organizing Control of Stochastic Systems, M. Dekker, NY, 1977.

[6] G. N. Saridis, "Foundations of the Theory of Intelligent Controls," Proc. of the IEEE Workshop on Intelligent Control, Eds. G. Saridis, A. Meystel, Troy, NY 1985, pp. 23-28.