Benchmarks for Mobile Robotics?

John Hallam & Gillian Hayes
Department of Artificial Intelligence
University of Edinburgh
5 Forrest Hill
Edinburgh EH1 2QL
SCOTLAND
{john,gmh}@dai.ed.ac.uk

5/9/97

Abstract

In many fields, efforts at standardisation and the definition of benchmark problems have resulted in greater coherence and focus. Researchers are able to compare their own work with that of others in a standard test scenario, and work is in principle easily replicated and exchanged.

In this position paper, we explore the possibilities, and their consequences, of introducing such benchmarking in the field of robotics research.

Introduction

In many fields, efforts at standardisation and the definition of benchmark problems have enabled researchers to share their results and insights and more easily to replicate each other's work. For instance, standard animal models in neuroscience, standard data sets in computer vision, standard laboratory equipment in molecular biology, have all contributed to the interchange of ideas, techniques and results.

The situation in robotics is rather different: while there are common concerns (for instance, many mobile robotics researchers study navigational skills — from the reactive, *e.g.* wall-following, obstacle avoidance, *etc.*, to more sophisticated skills such as localisation in Cartesian Metric Space or optimal path planning), work is conducted on different robotic platforms in differing environments with diverse software architectures and techniques. It is tempting to wonder whether a good set of benchmark problems is what we, as a field, need to enable us to progress as swiftly and effectively as those who pay our bills would wish.

Proc. Conference "Towards Intelligent Mobile Robots", TIMR 97, Manchester 1997.
Technical Report Series, Department of Computer Science, Manchester University, ISSN 1361 - 6161.
Report number UMCS-97-9-1. http://www.cs.man.ac.uk/csonly/cstechrep/titles97.html

In this position paper we consider some of the implications and prerequisites for useful standardisation of benchmarks. First, we think briefly about the nature of robotics — the problem of designing agents and the research agenda that arises from that view. Given that context, we move to the main argument of the paper and discuss the relevance and utility of benchmarks. Finally, we offer a short conclusion.

Robotics Research

It is now widely accepted that the behaviour of a system arises out of the interaction of three components: the task it is supposed to accomplish; the environment in which that task is to be tackled; and the system or agent whose job it is to accomplish the task in the environment. Varying any one of the three may change the final behaviour radically or minimally, and likewise the performance of the agent. This view of an agent, whose behaviour is determined by an interaction between its own processes and those of the environment in which it is placed, in the context of a task, places considerable demands on the designer of situated and embodied robotic systems, who must therefore construct a reliable interaction between a physical agent and a complex and potentially unmodellable environment in order to provide assurance that the required design task(s) will be achieved.

In general, there are four distinct kinds of solution a designer might deploy, assuming a given task to be performed, in response to the problems a robotic agent faces:

- *To engineer the environment.* Although this may seem a little strange, on reflection, this option is familiar to all of us. Human beings continually engineer their environment to make life easier for themselves. Our entire social infrastructure is based on an engineered environment, and we have real problems if the environment reminds us that we don't have it entirely under control: consider heavy snow in most British towns, for instance!

  The same idea can be used for intelligent agents. For example, we could considerably simplify the design of a street cleaning robot system by having it work during the early morning, between 2am and 4am say, when the environment is simplified by the low density of traffic on the roads; or we could eliminate the need for navigation in an office building by marking each significant place with a barcode corresponding to a reference on the architect's CAD drawings.

- *To change the agent's physical body.* Not applicable to software agents, this option is a real one for robotic agents. An appropriate choice of body can simplify the problems the designer faces in arranging for the agent to behave appropriately.

  For example, a robotic system designed to deliver mail in office buildings may face a problem with stairs. There are two obvious solutions to this problem not often considered: we could make the vehicle so large that it could not fall downstairs because it would not fit; or we could make it so small that falling downstairs would not matter. Either would avoid the need for environmental markings indicating dangerous stairs, or sensor systems and software to enable the agent to recognise and avoid the danger zones.

- *To change the kinds of action that agent can perform.* Similar to changing the agent's body, this option concerns the interface between the body and its command system. Consider an ordinary motor car. It is designed to go forwards and backwards, and to steer while moving. However, it cannot change direction while stationary, nor can it move sideways. This complicates parking (unacceptably so for some people). The problem is that the actions the car can perform are restricted: it might be useful to build cars whose steering and drive mechanisms would permit them to move in any direction from a standing start.

- *To provide control software to direct the agent's behaviour.* This is the first, and most obvious, solution to a programmer, computer scientist, or many workers in AI. One should solve the problems an agent encounters by providing it with 'intelligent' software, they suggest. As the examples above reveal, it is by no means the only option, and not necessarily the best.

Intelligent agents for any given task will typically face a number of problems of differing kinds, and solutions need to be found by the designer. It is usually possible to solve any given problem in several different ways. For example, the ability to deal with stairs can be implemented by labelling the stairs, by the choices of body size alluded to above, or by providing sensors and program for recognition of the danger and appropriate action. For each of these possible solutions there are consequences and trade-offs to take into account. A good designer will consider a wide variety of potential solutions, drawn from all of the classes outlined above, and will pick the most appropriate for the task at hand.

If this is the work of the robotic agent designer, then what is robotics research? We suggest it is the search for new solutions (or new *kinds* of solution) to known problems faced by robotic systems; an attempt to discover principled strategies and tactics for deploying the right kinds of solution to the problems encountered in design — that is, deisgn methodology (if one is an engineer); or (for the scientist) an attempt to identify the theoretical explanations and justifications of those solutions and rules in terms of some kind of Unified Theory of Agency.

Standardisation and Benchmarks.

The art of robotics, then, is the design of embodied agents that act in physical environments to achieve designer-specified (in reality, rather than the research laboratory, *customer-specified*) tasks. How might standardisation or benchmark definition fit into such a context?

Common Platforms?

One straightforward standardisation sometimes proposed is to adopt a common experimental platform. This has clear advantages of interoperability and ease of replication of others' work and a good example exists in the form of the Khepera, a tightly engineered miniature modular robot with precise movement and short range sensing now popular in a variety of mobile robotic experimental settings.

Consideration of the Khepera example illustrates the negative aspect of this form of standardisation, however: the platform has a strong influence on the kind of research tasks that can be tackled. The Khepera robot is essentially a table-top machine, requiring a surface smooth to within a millimetre or so, and requires a tether for experiments of long duration. The short range of its active sensors is also problematic. It has been used mostly for obstacle avoidance, simple navigation and manipulation experiments, and as a testbed for learning and evolutionary robotics; it is hard to imagine how algorithms developed on the Khepera might scale up to a machine working outdoors harvesting berries in the forest, for example.

Benchmark Tasks?

A second attractive possibility is the definition of a set of benchmark tasks for researchers to focus their experimentation on. An interesting parallel is with the use in neurophysiology of a small set of standard animal models (chosen in part for accessibility) on which a wide variety of studies can be carried out. Finding a set of such tasks is a daunting prospect, however: if too generic (*e.g.* navigation) the benefits of benchmarks are forfeited, while if too specific the problem of scaling up casts doubt on the utility of the standard. It is unclear (to us at least) whether a suitable set of tasks exists and what they might be.

Environmental Standardisation?

If standardising task or platform is problematic, standardising environments is challenging. It is possible to identify abstract properties of environments useful for classification (*e.g.* whether the environment is well-modelled by the Markov postulate; whether there are multiple agents present; whether the environment possesses its own dynamics or is essentially static; *etc.*) but the implications of these properties for behaviour and so for performance are not straightforward. For instance, an essentially Markovian environment will be experienced quite differently by agents whose differing sensoria allow them to perceive different amounts of environmental state. The problems are exacerbated in robotics where standardisation of environments may have to include even the details of the materials from which they are made, since in our experience variation in physical properties such as coefficient of friction can make dramatic differences to the behaviour of an agent.

Trade-Offs...

Looking back at the context in which robotic design takes place, it is unsurprising that the suggestions for standardisation offered above are deficient: each attempts to constrain one of the three determining components of interaction that enables agent performance while leaving the other two undetermined. Equally, an attempt to constrain all three would be unhelpful — a certain variety of research problem makes for an interesting field.

Constraining two of the three components is more promising. An interesting possibility is that the current popularity of robot contests and games may provide some insight into the definition of a set of challenging task-environment combinations that offer the potential to scale up to interesting applications.

Conclusion

Benchmarks and standardisation are certainly attractive but perhaps only seductive. They are not, however, a panacea for the methodological questions that our field raises. Certain standards and benchmarks may have great value, but they must be chosen with care and foresight. Standard robots, standard tasks to be accomplished, and standard environments in which agents operate are less attractive than they might at first seem.