European ROBOTICS Research Network EURON

IROS 2008 Nice

The Euron GEM Review guidelines: current state and perspectives

Fabio Bonsignorio
John Hallam
Angel P. del Pobil

- 'Look Ma, No Hands' sindrome?
- Replication of experiments
- Perfomance measure benchmarks to allow results comparison
- Needed to foster research advancement and enable practical application of research achievements

As the complexity of developed robotic and intelligent systems grows, it is more and more needed to define proper experimental approaches and benchmarking procedures.
Trustable benchmarks are needed in order to allow the comparison of the many research results in service robotics research end enable their industrial application.

Both replication and benchmarking are needed to foster a cumulative advancement of our knowledge of intelligent physical agents and even to correctly appreciate disruptive innovation in the science (NSF's CyberPhysicalSystem science subset???) and technology of robots.

Should we take inspiration from biology and medicine (just sample a typical Nature paper....)

If robotics aims to be serious science, serious attention must be paid
to experimental method.

What is an 'experiment' in robotics?

Replication&Falsification

As it is known K.Popper defined in a very tight way the requisites for a discipline to be considered 'scientific'.
In social science, management and economics exact repetition is often seen as a limit case, experiments that systematically vary one or more input parameters of a system under study to see whether its output parameters remain stable or change according to the expected model in a predictable way.

Replication&Falsification

Only when the model fails clearly in a number of variated experimental setup it is considered 'not replicable'. Nethertheless, as already noticed, all disciplines aiming to be considered 'scientific' incorporate a concept of experiment replication and a concept of 'falsifiation' of theory through experiments.

Replication&Falsification

There are different modulation of this concept, but whether we think we are in a cumulative phase in the development of a scientific field or in presence of a 'disruptive' creative paradigm shift, as somebody is claiming in nowadays robotics, a kind of widely accepted experimental methodlogy is needed in order to be able to ground the advancement of research on a shared quantitative language.

Replication&Falsification

A clinical trial protocol is the detailed written plan of a clinical experiment.
It may be inspiring looking at the US NCI guidelines for drafting a clinical trial protocol: the enphasis on signaling 'adverse events' , the definition of 'criteria for response assessment', the necessity of defining clearly principal and secondary hypotheses to be validated.
The statistical section of the protocol is asked to define how the data will be analyzed in relation to each of the objectives.

In particular it expects that an acceptable trial specify, with reference to the study objectives:

- Method of randomization and stratification
Total sample size justified for adequate testing of primary and secondary hypotheses
- Error levels (alpha and beta)
- Differences to be detected for comparative studies
- Size of the confidence interval of the estimates.

Replication&Falsification

It seems clear that in robotics the experimental methodology standards are currently in many cases weaker, and the sindrome 'it worked once, in my lab' could be more widespread than we may think.

Replication&Falsification

As already noticed, a limit to replication is given by the huge variability of robot machines.
Perhaps, following the biomedical analogy, we have to compare behaviors and performances of different 'animals'.
Anyway a wider capability to replicate research results is probably needed in order to allow a faster development of our field and to foster both cumulative progress and disruptive change.

Discussion

It seems that the bare replication of experimemts and the quantitative comparison of research results in robotics raise many challenging issues.
This is due to the variety of applications, tasks, mechanical structures, sensor sets, actuators, control system, software architectures, required levels of flexibility and autonomy, and so on.

Discussion

When we are dealing with Human Robot Interaction in everiday settings also human psicology is involved.
On the other end, there are many initiative trying to define proper standards.
There are benchmarks in some specific areas like visual servoing, SLAM, motion planning, but there is still a lot of work to do.

Discussion

Possibly we should identify a few limited and simpler tasks and related environments and develop benchmarks for those task that can be accepted and are by the community and then proceed extending the approach to more complex functions.
As told we should probably look to biology, medicine and 'soft' sciences for inspiration.

Discussion

In some experimental works ‘entropy measures’ on the ‘sensory-motor’ coordination of different ‘robotics’ equipment have shown that information metrics can be used to classify, at least, and to get an insight on (semi) autonomous robotics devices, which show an ‘emergent behavior’, while, in [Chatila,2006], entropy measures are used to rank enviroment complexity, with reference to the navigation task.

Discussion

An approach integrating task and environment complexities is proposed by Tardos et al.
HRI experimental research is sometime conducted by means of protocols deriving from psicology.

SIG on Good Experimental Methodology

- Chaired by Fabio Bonsignorio John Hallam and A. P. del Pobil
  - Benicassim, Spain, 9/1/2008, collocated with RISE'08
  - Genoa, Italy, 31/1/2008,
  - Zaragoza, Spain, 4/3/2008
  - Prague, 28/3/2008, as part of the EURON Annual meeting
  - Valencia, Spain, 30/4/2008

- Main result is a document entitled ***General Guidelines for Robotics Papers Involving Experiments***, available from the website and as an appendix in DR2.7
- With particular descriptions for the following domains:
  - SLAM
  - Mobile Robots' Motion Control
  - Obstacle Avoidance
  - Grasping
  - Visual Servoing
  - Autonomy/Cognitive Robotics

The Euron GEM Review guidelines: current state and perspectives

IROS'07 Workshop Program

- *Synthetic Approach to Cognitive Systems: A Perspective from Cognitive Robotics*, Kaz Kawamura
- *Benchmarking Urban 6D SLAM*, O. Wulf, A. Nuchter, J. Hertzberg, and B. Wagner
- *The Jacobs Test Arena for Security, Safety, and Rescue Robotics (SSRR),* A. Birk, K. Pathak, J., S. Schwertfeger, M. Pfingsthorn and H. Bulow
- *Towards Quantitative Comparisons of Robot Algorithms: Experiences with SLAM in Simulation and Real World Systems*, B. Balaguer, S. Carpin, S. Balakirsky
- *Reliability Testing for Embodied Autonomous Systems*, L. F. Gunderson and J. P. Gunderson
- *Advances in the Framework for Automatic Evaluation of Obstacle Avoidance Methods*, J.L. Jimenez I. Rañó, J. Minguez
- *Good Experimental Methodologies in Robotics: State of the Art and Perspectives*, F. P. Bonsignorio, J. Hallam, and A. P. del Pobil

OpenGEMForum 2008, Prague

25 March

- Open discussion about benchmarks and architectures for robustness and autonomy, H. Bruyninckx.
- Plenary discussion.
- Proposals for benchmarking SLAM, G. Fontana, M. Matteucci, J. Neira, D. Sorrenti.
- Motion Planning vs. Automated Planning in benchmarking, M. Reggiani, E. Pagello.
- Benchmarking mobile robots' motion, A. Marjovi, L. Marques.
- Plenary discussion.

26 March

- The Hydra-Shiva concept for GEM and Benchmarking in robotics, A Moshaiov.
- Plenary discussion
- Advances in the Framework for Automatic Evaluation of Obstacle Avoidance Methods, J Minguez.
- RoSta - A Brief View Over Benchmarking Activities In Service Robotics, K. Pfeifer.
- GEM and Benchmarking in robotics, where we are? Serious? Science?,
  F. Bonsignorio, J. Hallam, A. P. del Pobil.
- Plenary discussion.

The Euron GEM Review guidelines: current state and perspectives

European ROBOTICS Research Network EURON

IROS 2008
Nice

Workshop on Good Experimental Methodology & Benchmarks in Cognitive Robotics

- GEM and Benchmarking in robotics, F. Bonsignorio, J. Hallam, A. P. del Pobil.
- Quality Measures for Mapping: from Test Environments to Analysis Tools, M. Pfingsthorn, A. Birk.
- Benchmarking in the DEXMART project, G. Grünwald
- Experiences in evaluating human-robot interaction - COGNIRON, I. Lütkebohle
- Open Discussion

The Euron GEM Review guidelines: current state and perspectives

RSS'08 Workshop Program

- GEM
- Introduction, A.P. del Pobil
- The Jacobs Map Analysis Toolkit,
  I. Varsadan, A. Birk M. Pingsthorn, S. Schwertfeger, K. Pathak
- RobotStadium: Online Humanoid Robot Soccer Simulation Competition, O. Michel
- Can We Benchmark The Influence of Information-Processing Architectures on Intelligent Systems? N. Hawes, J. Wyatt
- The RAWSEEDS Proposal for Representation-Independent Benchmarking of SLAM,
  G. Fontana, M. Matteucci, D. G. Sorrenti
- A unified benchmark framework for autonomous Mobile robots and Vehicles Motion Algorithms (MoVeMA benchmarks), D. Calisi, L. Iocchi, D. Nardi
- Performance metric for vision based robot localization,
  E. Frontoni, A. Ascani, A. Mancini, P. Zingaretti
- The EURON GEM Review guidelines, F. Bonsignorio, J. Hallam, A. P. del Pobil
- Open discussion

The Euron GEM Review guidelines: current state and perspectives

General Guidelines for Robotics Papers Involving Experiments

1. Is it an experimental paper?
2. Are the system assumptions/hypotheses clear?
3. Are the evaluation criteria spelled out explicitly?
4. What is being measured and how?
5. Do the methods and measurements match the criteria?
6. Is there enough information to reproduce the work?
7. Do the results obtained give a fair and realistic picture of the system being studied?
8. Are the drawn conclusions precise and valid?

USEFUL LINKS

www.heronrobots.com/EuronGEMSig

www.robot.uji.es/benchmarks

Conclusion and future work

There is a widespread perception of the need of improving experimental prectices in robotics.
The Euron GEM Review Guidelines try to address these needs.
It is thought that proper and widely accepted replication procedures and performance benchmarks are needed to allow the cumulative progress of robotic science and technologies and even to assess the value of new disruptive ideas.

Thank you!

The Euron GEM Review guidelines: current state and perspectives

The Euron GEM Review guidelines: current state and perspectives