

Performance evaluation and benchmarking in Challenge 2 EU-funded activities

Cognitive Systems and Robotics

*European Robotics Forum 2011
08 April 2010*

Anne Bajart

*Unit E5 - Cognitive Systems, Interaction, Robotics
DG Information Society and Media
European Commission*

<http://www.cognitivesystems.eu>

Outline

- ⊙ FP7 Challenge 2 - Cognitive Systems and Robotics
- ⊙ Performance evaluation, Benchmarking, Standardisation:
 - ⊙ EU-funded effort
 - ⊙ An illustration in ongoing projects
- ⊙ Why Performance evaluation, Benchmarking and Standardisation?
- ⊙ Open questions
- ⊙ Brainstorming activities
 - ⊙ The EURON effort on benchmarking
 - ⊙ EUCogII - Challenges for cognitive systems
 - ⊙ The RoSta survey
- ⊙ What's next?
- ⊙ At an EU level

FP7 Work Programme

Challenge 2: Cognitive Systems and Robotics

It is important to be able to measure and compare progress towards the ambitious goals set under this Challenge.

Developing suitable benchmarks, conducting benchmarking exercises and supporting scenario-based competitions are therefore firmly placed on the agenda.

Challenge 2: Cognitive Systems and Robotics

FP7-ICT Call 7 Questions & Answers*

- © Q: Why is benchmarking important?
- © A: Benchmarking allows to **objectively evaluate** key system properties, depending on the particular R&D issues and application scenarios addressed. Each project is therefore expected to use reliable criteria for **assessing progress**, to make them **public** and, if possible, **compare** and contrast them with criteria proposed by other research groups.

*downloadable on <http://www.cognitivesystems.eu>

Robotics Benchmarking and Standardisation: EU-funded effort

- **EUropean RObotics research Network (EURON)**
 - <http://www.euron.org/>
 - Benchmarking Initiative <http://www.euron.org/activities/benchmarks/>
 - Research Roadmap <http://www.euron.org/activities/roadmap.html/>
- **Strategic Research Agenda - SRA**
(**CARE - EUropean RObotics Platform EUROP**)
<http://www.robotics-platform.eu>
 - Benchmarking and standardisation identified as a key requirement
- **euRobotics**
<http://www.eurobotics-project.eu/>
 - to give the academic world a chance to test the market-readiness of their technologies in scenarios selected by industry through competitions or Grand Challenges.
- **euCogll**
<http://www.eucognition.org/>
 - Towards the formulation and dissemination of « Challenges for artificial cognitive systems »

Robotics Benchmarking and Standardisation: EU-funded effort

- ***Robot Standards and Reference Architectures (RoSta)***
<http://www.robot-standards.eu/>
 - Action Plan for “benchmarking for mobile manipulation and service robots”
- ***Robotics Advancement through Web-publishing of Sensorial and Elaborated Extensive Data Sets (RAWSEEDS)***
<http://www.rawseeds.org>
 - Benchmarking toolkit for SLAM (data sets, benchmarks problems and solutions)
- ***Best Practice in Robotics (BRICS)***
<http://www.best-of-robotics.org/>
 - Harmonisation, Robot Software libraries, Methodologies, Showcases
 - Performance comparison of software and hardware
 - => Practicing benchmarking for robotics

Benchmarking and Standardisation: An illustration in ongoing projects

- **Smart Eyes: Attending and Recognizing Instances of Salient Events (SEARISE)** <http://www.searise.eu>
 - Performance evaluation of components against the existing approaches (common public image databases, comparison with human operator)
 - Benchmark datasets (available online)
- **DEXterous and autonomous dual-arm/hand robotic manipulation with SMART sensory-motor skills: A bridge from natural to artificial cognition (DEXMART)** <http://www.dexmart.eu/>
 - Definition of benchmarks and metrics suitable for performance evaluation of one and two armed/handed systems
- **Intelligent Surgical Robotics (I-SUR)** <http://www.i-sur.eu/> (to be confirmed)
 - To assess the feasibility of automation in minimally invasive surgery (for easy actions such as puncturing, cutting and suturing)
 - To demonstrate its value with realistic benchmarks and metrics

Why Benchmarking and Standardization?

- © To measure performance of systems
- © To test and evaluate in a reproducible way
- © To allow comparison of research results

USEFUL FOR

- © Scientific community: to focus efforts, to exchange results, to drive research and allow tangible progress
- © Industrial community: to assess quality, to meet users needs, to speed up development and testing time

Open questions

- How to evaluate a complex system?
- How to decompose into components or sub-components?
- How to define suitable metrics?
- How to reproduce an experiment?
- How to impose benchmarking as a scientific recognised valuable activity?
- How to benefit from standards without preventing innovation?

The EURON effort on benchmarking (2008 - ...)

© Research Benchmarks deliverable DR2.7 (2008)

Exhaustive lists and inventory of

- Benchmarking and Metrics Workshops
- Robotics competitions and challenges
- Benchmarks initiatives (inside/outside EU)

© Special Interest Group on Good Experiment Methodology

© GEM guidelines (2008)

towards “high quality reporting of replicable experimental work”

© Point of contact / room for discussion and collaboration

EUCogII

“Challenges for Cognitive Systems” (2011)

- A successful cognitive system is flexible (adaptive) and autonomous
- How to benchmark a full cognitive agent operating within an unknown environment?
 - Environmental complexity
 - Agent coping ability
- Systemic Challenges
(research lines: AI, perception, action selection, etc)
- Benchmark Challenges
(precise targets and performance levels)

The RoSta survey (2009)

- Robotics Technologies to benchmark: navigation, grasping, reliability (degree of failure), autonomy, specific tasks involving various technologies (scenarios)
- Beneficial standards: non-profit, clear specifications, allow comparability of systems and components
- Contributors: academic researchers, industrial world, end-users
- Benchmarking culture: centralised certifying body, mandatory scientific activity



What next?

- EU-funded research is an asset to gather efforts, define metrics and develop benchmarks
- Standardization will help
- Open-source technologies are a starting point
- Think global, not local: your project results will benefit the whole community
- How to impose a benchmarking culture?



FP7 - ICT Call 9 - 2.1 Cognitive Systems and Robotics

Publication: 18/01/2012 (TBC)

Deadline :17/04/2012 (TBC)

- © Target (c): Gearing up and accelerating cross-fertilisation between academic and industrial robotics research

synergies between respective research agendas through joint industrially-relevant scenarios, shared research infrastructures;

joint small- to medium-scale experimentation with industrial platforms and

implementation of comparative performance evaluation methodologies and tools.



Thank you for attention

Anne BAJART
Project Officer
European Commission - DG INFSO
Unit E5 - Cognitive Systems, Interaction, Robotics
e-mail: anne.bajart@ec.europa.eu



FP7 - ICT Call 9 - 2.1 Cognitive Systems and Robotics

© Target (e): Speeding up progress towards smarter robots through targeted competitions

- based on suitably evolving reference scenarios
- focused on capabilities
- involving relevant stakeholders

events, dissemination and public awareness measures



At a EU level

As a EU-funded project, you have a RESPONSIBILITY

- © Share your knowledge
- © Help building a better science, a more competitive European industry
- © Exchange your data, your results, your best practices
- © Build a network
- © Think benchmarking and standards
- © Go one step further and submit a project proposal about benchmarking and standardization