RoboCup@Home — a competition as a testbed for domestic service robots

Dirk Holz[1], Luca Iocchi[2], Javier Ruiz-del-Solar[3], Komei Sugiura[4], and Tijn van der Zant[5]

[1] University of Bonn, Germany
[2] Sapienza University of Rome, Italy
[3] Universidad de Chile, Chile
[4] National Institute of Information and Communications Technology, Japan
[5] University of Groningen, The Netherlands

1 Introduction

Scientific competitions significantly increased their scope and diffusion in the last years and nowadays many research areas have some type of competition. They provide frameworks where research groups can compare their approaches and also give the opportunity of defining standard test beds for solving specific problems. Moreover, the tasks and the results of competitions are often used in scientific papers to compare new approaches with respect to existing ones.

Some of these competitions are run periodically (e.g., every year), changing tests over time for addressing more difficult problems and enlarging their variety. In order to guide such a progress and have an impact on how solutions are developed, it is important to evaluate the performance every year and to define new tasks together with the scoring that will determine the rank of the participants.

RoboCup@Home is a competition where domestic and service robots are performing several tasks in a home environment, interacting with people and with the environment in a natural way. Each test requires a combination of different functionalities and the score is related to the accomplishment of the task. In this extended abstract (and talk), we will describe how competition results are analyzed in RoboCup@Home and how the results of this evaluation drive the development of future competitions.

2 The RoboCup@Home Competition

RoboCup@Home competitions include an annual international event within RoboCup and several regional events in Europe, Japan, USA, etc. The competition is run in a realistic setting where an apartment with different functional rooms and typical furniture and objects is realized. This environment (size, material of the floor, colors of the wall, kinds of furniture and objects, etc.) are not known in advance which ensures the development of general solutions. Moreover, some tests are performed outside this environment, in public crowded places.

The competition is formed by about ten tests, a Technical Challenge and the finals. The main functionalities required in the tests are mapping and navigation, person tracking and recognition, object recognition and manipulation, speech and gesture recognition as well as general cognition abilities. Each test requires a combination of some of these functionalities. For example, in *Follow me*, the robot has to follow a previously unknown person through a crowded previously unknown area, passing different obstacles including getting hindered from following and entering an elevator in order to reach a location on a different floor. Another test is *Cocktail Party* in which the robot has to find persons in an apartment, and take and deliver beverage orders. The *General Purpose Service Robot* test (GPSR) focuses on the ability of the robot to understand its goals and reason on them. The task is not specified beforehand, but generated randomly and it is given to the robot through a spoken command. The robot has to understand the desired goal and plan actions accordingly. In this test, all functionalities may be required, but in particular the robot must demonstrate the ability of understanding the current situation, the user request and faults such as commands containing erroneous or missing information. In case errors occur, the robot needs to report these and find alternative solutions. In case of missing information, the robot has to initiate a dialog with the user to obtain the missing bits. A more detailed description of all tests in the competition is presented in [1] as well as in the RoboCup@Home rulebooks[6].

Each test requires a proper integration of a subset of functionalities. We divide tests into multiple phases (or sub-tasks) and each phase, when accomplished, provides the team with a score. The total score of the test is thus given by the sum of the scores of all the accomplished phases. These partial scores are specified in advance and allow to evaluate the complexity of the underlying problem. We associate to each test phase a set of functionalities that are required to achieve it. When a phase is successfully accomplished, we can state that all involved functionalities have been successfully implemented.

3 Score System and Analysis

In order to measure progress in the competition, let us consider a test T, divided in n phases $(p_1, ..., p_n)$ and denote with $F(p_i)$ the set of functionalities associated to each phase p_i. Then, we define w_{f,p_i} as the weight of the functionality $f \in F(p_i)$ in p_i, that can correspond to the probability $p(f|p_i)$ that f is the cause of failure of p_i. For each phase that a team accomplishes, the score S_{f,p_i} is assigned to each functionality f associated to the phase according to the weights. Then, all the scores for each functionality achieved by a team in all the tests are summed and this value is normalized over the maximum value of score that a team could have achieved on this functionality. This normalized value $\bar{S}_f^{\delta}$ is thus the percentage of score achieved by a team δ in the functionality f during all the tests. Table 1 shows the maximum percentage score $\hat{S}_f^{\Delta}$ achieved in every functionality by the teams over the past six competitions. Results like these are

[6] `www.robocupathome.org/rules`

Functionality	2008 [%]	2009 [%]	2010 [%]	2011 [%]	2012 [%]	2013 [%]
Navigation	40	47	33	61	52	27
Mapping	100	100	21	33	10	26
Person Recognition	32	69	57	48	62	51
Person Tracking	100	100	100	100	62	51
Object Recognition	29	39	6	35	56	37
Object Manipulation	3	48	29	49	73	45
Speech Recognition	87	89	50	76	90	46
Gesture Recognition	0	0	62	100	88	44
Cognition	-	-	17	68	32	36
Average	**48.9**	**61.5**	**41.6**	**63.3**	**58.2**	**40.3**
Standard dev.	**41.3**	**34.8**	**22.1**	**25.1**	**26.8**	**9.4**

Table 1. Achieved max. scores for the desired abilities.

very important since they show how the competition is progressing and possibly how to modify the rules in order to drive the development of the competitions. Some general rules that have been applied in RoboCup@Home are:

1. if a functionality has high score (e.g., Mapping in 2009), then increase its difficulty in the next year;
2. if a functionality has low score (e.g., Object Recognition in 2010), keep the difficulty unchanged;
3. if a functionality is not developed at all (e.g., Gesture Recognition in 2009), make it mandatory and increase the value and the phases in which it is used.

The main goal of this analysis and of the consequent measures is to keep a reasonable level of difficulty over the years and to balance the development on all the functionalities. Moreover, in development of integrated research in AI and robotics, a good balance of all the functionalities is required. The low standard deviation obtained over the years in RoboCup@Home, as shown in the table above, demonstrate the efforts of the competition organizers and of the teams in developing systems that can properly integrate many functionalities.

Summarizing, by monitoring these results over the years, we can measure progress and success in task accomplishment and further drive the competition into directions considered interesting.

References

1. Holz, D., Iocchi, L., van der Zant, T.: Benchmarking intelligent service robots through scientific competitions: the robocup@home approach. In: Proceedings of the AAAI Spring Symposium Designing Intelligent Robots: Reintegrating AI II. (2013)

Fig. 1. Teams and robots at the RoboCup@Home competition 2013 in Eindhoven, Netherlands.