Evaluating the use of robots to enlarge AAL services

Cecilio Angulo*, Sammy Pfeifer†, Ricardo Tellez‡ and Guillem Alenyà‡§

*Department of Automatic Control, Technical University of Catalonia, Pau Gargallo 5, 08028 Barcelona, Spain
†PAL Robotics, Pujades 77, 08005 Barcelona, Spain
‡Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain
§Corresponding author: galenya@iri.upc.edu

Abstract—We introduce robots as a tools to enhance Ambient Assisted Living (AAL) services. Robots are a unique opportunity to create new systems to cooperate in reaching better living conditions. Robots offer the possibility of richer interaction with humans, and can perform actions to actively change the environment. The current state-of-art includes skills in various areas, including advanced interaction (natural language, visual attention, object recognition, intention learning), navigation (map learning, obstacle avoidance), manipulation (grasping, use of tools), and cognitive architectures to handle highly unpredictable environments. From our experience in several robotics projects and principally in the RoboCup@Home competition, a new set of evaluation methods is proposed to assess the maturity of the required skills. Such comparison should ideally enable the abstraction from the particular robotic platform and concentrate on the easy comparison of skills. The validity of that low-level skills can be then scaled to more complex tasks, that are composed by several skills. Our conclusion is that effective evaluation methods can be designed with the objective of enabling robots to enlarge AAL services.

I. INTRODUCTION

The domestic environment is a major growth area and a prime site for the development of Ambient Intelligence ('smart home'). For example, people aged 65 and older are the fastest growing segment of the US population; by 2030 almost 9 million Americans will be over age 85, and 72.7 million over age 65 [41]. Moreover, over 20% of people 85 and over have a limited capacity for independent living, with the result that they require continuous monitoring and daily care [5]. The creation of a secure, unobtrusive, adaptable environment for monitoring health and encouraging healthy behaviour will be vital to the delivery of health care in the future.

According to these goals, research on Ambient Assisted Living (AAL) aims to develop solutions that increase independence for dependent and older people, and also for their formal or informal carers. It is an important point to address these solutions providing ways for older people to remain active and connected to society. As a hot research topic, funding activities like AALJP (Ambient Assisted Living Joint Programme) aim to create better condition of life for the older adults and to strengthen the industrial opportunities in Europe through

the use of information and communication technology (ICT) [46]; international workshops and journals describe progress on home-care, smart environments and other research topics for AAL such as human-computer interaction at assistive environments, semantic modelling for realizing reactive and proactive assistive environments, sensing, monitoring [9], [39]; and some competitions exist evaluating AAL systems through competitive benchmarking [8], [14].

EvAAL is a competition that aims at establishing benchmarks and evaluation metrics for comparing Ambient Assisted Living solutions [14]. Since 2011 international competitions have been organised on specific aspects of AAL systems, with the long-term goal of evaluating complete AAL systems. Three tracks are currently being deployed, dedicated to different aspects of indoor localisation and activity recognition for AAL. Since research on robotics is progressively focusing its attention to the design and development of companion robots to help seniors and disabled people living independent and safe in their own homes [12], there was a track about "Companion Robots for AAL" in the 2013 EvAAL edition [13]. This track aims at defining benchmarks for the integration of such robots with intelligent environments that provide user localisation and situation detection. The purpose is facilitating human-robot interaction in AAL environments.

At the same time, the robotics community is also approaching smart environments from the robotic perspective in the form of service and social robots. Hence, the robotics competition RoboCup [11] contains a track named RoboCup@Home with the aim to develop service and assistance robots to help out in domestic environment. A set of benchmark tests is used to evaluate the robots' abilities and performance in a realistic non-standardised home environment setting. Focus lies on domains like human-robot-interaction and cooperation, navigation and mapping in dynamic environments, computer vision and object recognition under natural light conditions, object manipulation, standardisation and system integration [53], [57].

A similar initiative, the RoCKIn project, is currently being designed as an European project including robot competitions acting as a catalyst for smarter, more dependable robots, built upon the principles of challenge-driven innovation laid down by RoboCup. In particular, the RoCKIn@Home challenge focuses on domestic service robots: robots that have enhanced

networking and cognitive abilities and are able to perform socially useful tasks such as supporting the impaired and the elderly [30], [45].

Both communities, researchers on AAL working with robots and researchers in robotics working in companions / social / assistive robots can benefit each other from evaluation experiences in benchmarking. However, from our experience in several robotics projects and principally in the RoboCup@Home competition, it can be affirmed that it is still an open problem to determine valid and accepted evaluation metrics in benchmarks comparing robots enlarging AAL services [20], [51]. Hence, from the point of view of ambient intelligence, competitions like EvAAL are now considering robots for human behaviour understanding from a high-level skills perspectives. Besides, a large experience exists in RoboCup@Home about using robots in a number of challenges, some of them very related with EvAAL, so they can be extrapolated to. In particular, measurements used in RoboCup@Home can be also extrapolated. However, some critics appeared in precedent editions of this robotic competition about how measures were designed, too specifics on low-level robotic skills, so, when exporting to EvAAL, a feedback loop should be considered to improve these measures as well as adapt them to this new environment.

It should be also mentioned the PerMIS (Performance Metrics for Intelligent Systems) workshop series, started in 2000 [34], aimed towards defining measures and methodologies of evaluating performance of intelligent systems. This workshop focuses on general applications of performance measures to practical problems in commercial, industrial, homeland security, and military applications. Therefore, some valid conclusions could be obtained from this initiative, but its starting perspective is approaching the problem from the general concept to the particular domain, in the inverse sense to our approach.

In short, our main contribution in this work is to propose a new set of evaluation methods to assess the maturity of the required robotic skills when solving challenging AAL tasks. Such comparison should ideally enable the abstraction from the particular robotic platform and concentrate on the easy comparison of skills. The validity of that low-level skills can be then scaled to more complex tasks, that are composed by several skills.

The rest of the paper is organized as follows: benchmark tests in RoboCup@Home that could be easily translated to the EvAAL and related AAL challenges are introduced in Section 2; in particular, using robots to enlarge AAL services is defended in two scenarios, at home and at supermarket. The proposed measures and evaluation methods are introduced in Section 3, as well as it is explained how they can be translated from robotics competitions (RoCup@Home) to AAL ones (EvAAL); finally, some conclusions and future research lines are offered.

II. Robotics challenges into AAL competitions

Benchmarking is important in research and technical developing because it allows to objectively evaluate key system properties, depending on the particular issues and application scenarios addressed. Reliable criteria should be used for assessing research progress, they should be public and, if possible, compared and contrasted with criteria proposed by other initiatives. Competitions stimulate innovation in a compelling way by providing researchers and developers with realistic mock up scenarios to test and compare their systems [31].

Challenges in the RoboCup@Home league and similar robotics competitions aim to develop service and assistive robot technology with high relevance for future personal domestic applications. RoboCup@Home is the largest international annual competition for autonomous service robots. A set of benchmark tests is used to evaluate the robots' abilities and performance in a realistic non-standardized home environment setting (see Fig. 1).

A certain number of benchmark tests exist in a first stage in order to check robots' structure and their basic abilities, as well as open challenges designed for the competitors in order to demonstrate the robot's technical skills. The list of very precisely defined benchmark tests challenging robots' performance is:

1) Enduring General Purpose Service Robot. This test focuses on robot reasoning and robustness.
2) Restaurant. The focus of this test is mobile manipulation in a real environment, such as a restaurant or shopping mall.
3) Emergency situation: home accident. Automatic detection of events to improve safety and security in the house in case of emergencies.
4) Follow me. The robot has to safely follow an unknown person through a public space.
5) Cocktail Party. The robot has to learn and recognize previously unknown persons, and deliver food orders.

It can be checked that the first two benchmark tests are very robot-oriented (manipulation, cognition, robustness) and the third test is usually home-oriented (home accident), with low human-level considerations. Hence, attending to our experience, the most suitable ones in AAL competitions, having strongly in mind interaction with humans, are:

Follow me. This test focuses on tracking and recognizing a previously unknown person, basic interaction and signaling capabilities, and safe navigation in unknown environments and narrow spaces with other people walking around or blocking the way.

Cocktail party. This test focuses on the following skills: human detection and recognition, manipulation, safe navigation and human-robot interaction with unknown persons. Figure 2 shows the 2013 score sheet for this test.

These two tests work as a perfect example on the integration of robot skills with what is done in AAL competitions like

Fig. 1. Scenario used in the RoboCup@Home competition for most of the challenges. It reproduces a common apartment, with lower walls to see the evolution of the robots along the tests. The scenario is composed by the entrance, the kitchen, a rest room and a living room with real furniture.

EvAAL. The same environment that can be found on an AAL environment, like the CIAMI living lab at the Technologies for Health and Wellbeing installations, in Paterna (Valencia, Spain) [35], is the one that is used for the RoboCup@Home (see Fig. 1). The main difference is that in the robotic competition there are no external devices in the house. The only sensory system is the robot. Additionally, the robot has the capacity to act on the environment, grasping, delivering or even requesting information from users around.

The unavailability of external sensor devices is a point to be solved in robotics challenges in order to converge to AAL competitions. Hence, for instance, looking at the preliminary rules of the new 2014 RoCKIn@Home competition[1], it can be checked that external devices like sensors and actuators are currently being considered: a camera-based intercom, lamps and shutters on the bedroom window are accessible and controllable via network, and networkable cameras in the kitchen.

This convergence between both worlds should walk forward also from the AAL competitions side. Unfortunately, none of the three tracks in the 2014 EvAAL competition is considering the use of service / assistant robots like a main solution for human behaviour understanding. In this form, the RoCKIn initiative is taking advantage in the AAL-robotics convergence. Moreover, from the 2013 EvAAL Companion Robot Competition, only one activity called "Autonomous movement started by a user request" is challenging enough compared with robotic competitions.

Interestingly, however, usability and acceptability are features to be specifically evaluated in the EvAAL competition, although in a not precise form, and they are not explicitly

[1]http://rockinrobotchallenge.eu/rockin_home_nutshell.pdf

considered in robotics benchmark tests. This is a point of improvement for both communities when evaluating domestic robots in AAL activities [4].

A. Using robots in AAL. Enlarging AAL services from the robotic approach

In the same form that recognition / understanding of human behaviours in AAL activities grows from simple tasks (walking, sitting down, standing up) to more elaborated ones (cooking, watching TV, having a shower) [1], [32], ideally robotic benchmarks should be also designed for each of the different modules from basic skills to more complex activities.

To obtain robotic systems with useful skills for real domestic applications, robots should be autonomous and with enough mobility. Activities with remote-controlled robots should be avoided as evaluated challenges, but maybe for the robot checking initial stage.

1) Low-level robotic skills in AAL: Hence, desired low-level abilities to be investigated and challenged are:

a) Navigation in dynamic environments.: Robots have traditionally navigated in static environments with few simple obstacles. However, real environments like homes and offices require to take into account complex scenarios, with people moving and possibly interacting with the robot (see Fig. 3). Moreover, bigger spaces with lots of people have to be considered, like shopping malls and airports. Safe navigation in crowded scenarios is still a subject of research, and most of the work has been only completed in simulation [25], [49], [50], [52]. Additionally, a map, built off-line, is required.

b) Fast and easy calibration and setup.: Usually several calibration procedures and arrangements have to be performed before a robot is able to perform tasks. Some examples of such

The maximum time for this test is *10 minutes.*

Action	Score
Taking the orders	
Detecting the calling persons	*3 × 150*
Understanding and repeating the correct person and drink names (half points are given if either of the names is incorrectly understood)	*3 × 100*
Grasping drinks	
Grasping an ordered drink (and successfully lifting it up to at least *5 cm* for more than *10 seconds*)	*3 × 200*
Delivering drinks	
Delivering the *correct ordered drink*	*3 × 200*
Leaving the arena	
Autonomously leaving the arena within the time limit	*50*
Penalties	
Talking to the wrong person (using wrong name)	*-50*
Talking to something that is not a person	*-150*
Special penalties & bonuses	
Not attending (see sec. 3.7.1)	*-500*
Using start button (see sec. 3.3.5)	*-100*
Outstanding performance (see sec. 3.7.3)	*200*
Total score (excluding penalties and bonuses)	*2000*

Fig. 2. An example of score sheet used for the Cocktail Party test in the RoboCup@Home league. Observe that if the verbal interaction fails (taking the orders) the test cannot continue.

Fig. 3. Robots need to able to navigate in dynamic environments with people in a safe manner, taking into account the possible interactions.

procedures are: mapping the environment, put semantic labels to the map and learn the appearance of the objects. The long-term objective of the competitions should be to have robots ready to function out of the box. This is a key objective for both, AAL and robotic communities. Score sheets in AAL

competitions explicitly take into account integrability and installation complexity, whereas for the robotic competitions is not yet a relevant topic as it can be deduced from its wiki page[2].
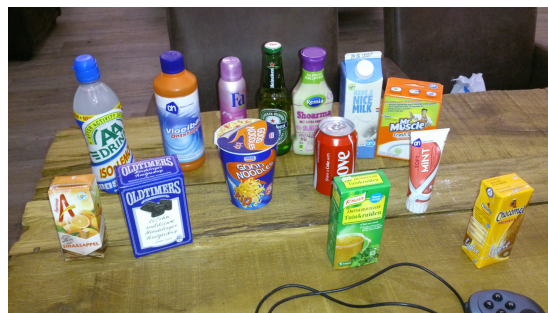

Fig. 4. Usual objects in the competitions. It can be observed that generally all the objects are geometrically simple and highly textured.

c) Object recognition.: This ability includes the identification of the different objects in the scene, as well as an off-line learning step to gather information about the texture, the geometry, and a semantic label. The recognition algorithms should be robust to changes in the orientation and position of the object, and also to the illumination conditions. The kind of objects used in competitions is usually relatively easy (see

[2]robocup.rwth-aachen.de/athomewiki/index.php/Publications

Fig. 4). The challenge is to detect one object under severe changes in the view-point, as one object can be placed on a table, on an elevated shelf, or even on the floor. There are some initiatives to provide generic descriptions of objects in the cloud that can be downloaded and used by robots [55].

d) Object manipulation.: In some tests, robots are required to grasp objects and transport them. It is mandatory that the motion of the arm is free of collisions, for example with other near objects or with the table or shelf where it is standing. Also it is required that the grasping is stable, that means that the object should be firmly attached to the hand. Textiles and other deformable objects are common in everyday environments, but still not considered in competitions as deformable object manipulation is challenging (see Fig. 5). Observe that the pick-and-place action requires grasping, transporting and releasing. In competitions usually grasping is evaluated separately from releasing (see the score sheet in Fig. 2), and both actions grant points.

Fig. 5. The REEM robot grasping a textile. Textiles and other deformable objects are common in everyday environments, but still not considered in competitions as deformable object manipulation is challenging.

e) Detection and Recognition of Humans.: A specific human recognition module is required to learn the appearance of humans and associate it with a given name. The objective is that the robot recognises people he has already seen, and correctly identifies when a person is new [26], [54]. This is a direct connection with AAL searched skills, hence robots act like sensor devices in the home environment, as well as processing elements.

f) Speech recognition.: Speech is the most natural form of interaction with the robot. However in competitions only a small subset of words are required to solve tests. For example, the names of the people are chosen from a small set of predefined names. The main challenge in real crowded scenarios is to handle the high level of ambient noise [58]. A specially challenging topic is solving this issue without particular changes in the physical robot, like external microphones.

g) Gesture recognition.: An additional way of communication with the robot is gestures [37]. In competitions, very simple gestures are considered, like waving to recall robots attention, or stop gestures. The challenge is to combine gestures with other input methods, like verbal input, to obtain better human robot interaction experiences [15]. For example, using combined input methods the command of pointing one object to ask the robot for a grasping operation can be disambiguated from an action to remember the position of the object.

2) High-level robotic skills in AAL: Next, higher-level skills that require the combination of some basic skills are listed:

a) Natural human-robot interaction.: Natural human robot interaction (HRI) is a hot topic which is being promoted in the robotics community. The objective is to develop systems capable of naturally interacting with people with everyday expressions and motions in the two directions: from the human to the robot, and from the robot to the human. An important point is that users are not required to have deep knowledge about the robot operation. The challenge in the human-to-robot direction is to combine several basic skills to obtain more elaborated percepts, like discovering human intentions [29], [44], [56] (for example, while communicating by voice, gestures or shared attention). The challenge in the robot-to-human direction is safety, that has to be embedded and granted in every robot action [22] (for example, when crossing with humans in a corridor).

b) Robot cognition.: The effective use of robots in everyday life situations is still far from being feasible. The main issue continues to be the great variety of unpredictable situations a robot has to face. Classical programming cannot provide a suitable solution to the general situation, so the trend in competitions has been to start with very simple tests, and slowly increase its complexity. The use of cognitive architectures is the preferred solution to handle more general tests. Recently, the deterministic classical cognitive architectures to take decisions have been applied to competitions [42]. However, uncertainty is a hard problem in real scenarios, and new paradigms should be developed [36]. The big issue with explicitly taking into account uncertainty is the poor scalability of the solutions. Some real situations to evaluate robot cognition includes planning actions requiring several nested steps, for instance requesting the robot to bring a drink from the kitchen, to bring a person from the entrance, to identify and solve an emergency situation.

c) Ambient intelligence.: The inclusion of ambient intelligence to the current competitions is still very limited. The robotic competitions consider that the sensors are all embarked on the robot. This is a robot-centric point of view, where the robot can take profit of some external sensors, for example a fire alarm on the house. The real objective should be to

communicate with surrounding devices, getting information from the Internet, and consider the robot as an additional device of a complete AAL system.

B. Using robots in AAL. Two possible scenarios

The objective of this section is to propose two different scenarios for testing and benchmarking robots integrated in ambient assisted living systems. The main goal is to propose tests that take advantage of the differential capabilities of a robot when compared with standard AAL systems. The most important characteristics that can be exploded are:

Integrability and installation complexity. When designing an AAL system, there would exist a balance between the time of deployment versus the accuracy that the system can obtain. Besides, the capability of having movable sensors is clearly valuable, as it has been already implemented in the 2014 RoCKIn@Home competition rules.

Physical interaction with the environment. Robots can perform actions in the environment and it is a unique opportunity to enlarge AAL capabilities. For example, acting on home appliances (ordinary lights switches, doors), grasping and moving things (bringing objects to the user), and using human tools (cooking, ironing, folding clothes).

Friendly interaction with the system. A AAL system should provide an easy interaction with the user. A robot offers a unique opportunity for the designers to embed the interaction with the AAL system though a robot in a natural manner. Additionally, a robot can easily provide feedback about the completion of the demanded tasks and reports about the status of the system.

One of the challenges when joining efforts between robotics and AAL communities is in middleware frameworks. The robotics community has adopted ROS (Robot Operating System) [43] as a standard de-facto and it is being used to effectively communicate processes in distributed systems, as well as easily integrate very different libraries in a common framework. Efforts for standardization have been also made in the AAL community. For example, the UniversAAL project [47], and the HOMER AAL components [23].

There have been some attempts to bring robots to AAL systems. In the framework of the DOMEO project, two different robots were developed, one as a walking assistant and another as a cognitive companion [19]. Another example is in the context of the FLORENCE project, where a simple tele-presence robot is developed to support eldery people at home [33]. However, there is no standardization of procedures and there is not a formal comparison and benchmarking regarding robots in AAL.

1) The robot assistant: One of the most investigated applications for service robots is as personal assistant. Both RoboCup@Home and RockIn@Home competition include tests in that direction. For example, a person asks to the robot to bring a drink from the kitchen, or a elderly person asks the robot to look for her glasses. These tests are very interesting for both, the specialists and the general public.

From the AAL perspective, the environment play a key role in the proposed test. In the 2013 edition of the EvAAL competition there was a demonstration track that included a robot. An important topic in EvAAL is people localization and activity tracking, and the proposed test might take advantage of such technologies.

The focus of this test can be oriented towards the next scenarios:

- Tele-assistance: the robot can be used to make tele-assistance services more effective. Taking the advantage of AAL localization, when an incoming call is at home, the robot can ask to the system where the person is located. The system asks the assistant to wait until he reaches the person.
- Emergency: Taking advantage of AAL activity recognition, the system can recognize an emergency situation, for example the person falls on the floor. The system asks the robot to approach, and can connect to the emergency system where a doctor can start the diagnose. When the emergency services arrive, the robot can open the entrance door.

The general limitations to the environment that should be added in the rule book of the competition are mainly two. The first one, is that the space has to be designed as an open space with no doors, except for the main entrance door. Currently robots are able of opening and closing doors, but this is an operation that takes a long execution time [38], [6]. The second one is to consider a planar floor, with soft slopes but no stairs. Most of the indoor robots are designed with a wheeled base and are not capable of traversing stairs. Studies with robots able to deal with stairs would be of interest [17], [18] in the robot assistant domain.

2) The shopper robot: In contrast to domestic environment, this is a test in an open and less controlled environment. In this sense, EvAAL proposes some tests in big spaces related to people localization using RF-based technology.

In this test, the environment is a supermarket. The robot should follow the user in a safe manner, while walks around other people and helps him / her in the shopping task. The robot helps to carry goods, answers questions like the comparison of prices, the availability of goods, and even about the list of ingredients of a recipe [24], [28]. The proposed scenarios differ a bit if a personal robot is considered (the owner goes to the supermarket with his own robot), or if a marketing robot is considered (the robots are on place waiting for customers). Navigation is also a challenge, as supermarket represents a crowded environment. In the RoboCup@Home competition and example of difficult navigation is benchmarked. It consists in a robot following a person that has to get in and out of an elevator (see Fig. 6).

The focus of this test can be oriented towards the next scenarios:

- Guiding the customers in the supermarket. This task can be oriented as assistance for elderly people, but also to the general public [28]. The AAL system can retrieve the

Fig. 6. One of the navigation challenges in the RoboCup@Home competition is to follow a person that enters into an elevator.

shopping list from the user's mobile device, and plan the best shopping experience based on constraints like time, distance, avoidance of lines... The robot can be used to carry the goods or push the shopping cart. Extensions to other similar large environments like museums are fairly direct and analysed in the literature for a long time [7], [10], [16].

- Automatic shopping bill. While the customer is selecting goods and putting them to the cart the robot can perform on-line the bill. This is an useful information for the user to monitor the spent, but also alleviates lines on the counter.
- Suggesting to the customers healthy meals.
- Promoting products. It is worth to notice that marketing robots is one of the areas promoted by EU in the Horizon2020 programme because it is considered that can be in real life in a short time. Robots could also promote themselves [21].
- Safe navigation. The AAL system keeps track of free and congested areas, and can provide this information to the robot to choose safer navigation plans.
- The AAL system keeps track of the position of the kids using RF-based systems.

The very general limitation to be introduced here is that real crowded environments can not be handled by current robot, as usually safety measures cause that the robot cannot move. Thus, the idea of using environmental monitoring to help the robot in navigation is relevant. Another limitation is in the set of different object that robots can consider at the same time. One promising alternative is to use the ideas of cloud robotics [27], [48], where robot can retrieve missing information from services in the cloud.

III. BENCHMARKING, MEASURES AND EVALUATION

In this section, some rules are provided to create a proper set of common benchmarking tests for competitions in the style of EvAAL [3], RoboCup@Home and RoCKIn@Home, based in our experience on robotics competitions and AAL research.

A. Lessons learned

We have a large experience in using humanoid robots for service at home. The REEM series of robots have participated at several editions of the RoboCup@Home (2007, 2012, 2013, 2014) (see figure 7) [2], [59], and several RoboCup German Open (2007, 2013). We are also collaborators and providers of the REEM platform at the RoCKIn@Home 2014 competition for teams without a robot.

Fig. 7. REEM robot delivers inscription form of the REEM@IRI team to jury during the RoboCup@Home 2013 competition.

Besides, we are participating in technical committees of AAL competitions, workshops, and participants in EU projects on AAL.

From that experience, it has been identified that in order to count on a good benchmarking measure, the following criteria should be included:

a) 1 - Agnostic implementation.: The first observed fact is that in order to apply a measure to any competition (or even to score a robotic product), the measure must be independent of the implementation. This means that it must avoid to include in its score calculation some elements that eventually

could benefit implementations specifically designed to solve the challenging test. For example, even if the same home environment can be found on an AAL environment as well as at the RoboCup@Home competition, that is the home, both approaches differ qualitatively in their solutions to provide a comfortable assisted environment to the users. Meanwhile the former relies on external devices distributed along the house (including sensors and actuators), the later relies on a mobile robot that includes in itself all sensory and actuation. Since the interest is in solving the general problem "to help at home", the scoring system shouldn't reward more any of those implementations, as well as to any other combination of both.

b) 2 - General over specific.: The aim of the competitions is for general purpose solutions, not for very specific ones (they aim a system that does it all)

Since it is very difficult to do a general test of everything, competitions have several tests each one focused on solving a particular tasks (see section 2). For example, there is a task that requires the robot to follow a person, and another that requires the robot to bring something to the user. The goal is to have a set of tests that an AAL system must pass in order to be marked as general.

However, we observed that some teams design specific solutions for a few tests (scoring very good at them), and forget the rest (hence, scoring very bad at the rest of tests). They specialize in a few tasks (may be just speech recognition and face recognition) winning points on every test by doing only those things. This misbehaviour leads to strange situations like a robot scoring in the top-three in the competition but being unable to do a single grasping behaviour (this type of skill is something very basic for a robot that should manage on a home environment). However, such a robot is able to obtain many points using low level skills along the competition. Having obtained many points at the few tests, they can over perform in terms of points to teams that have built more general solutions.

c) 3 - Progressive score.: Having a score of the type "all or nothing" is a bad idea. The score system cannot only reward 100% completed tests, it should provide intermediate points as the test is 1/4th, 1/2th and 3/4ths, otherwise there could exist a certain difficult step that will put many teams at the same score, even if some achieved a little more than others. Instead, a graded score is preferred because it better reflects the different status of the different solutions.

This can be done by introducing several points for intermediate resolution of the whole task. The problem is finding how to reward and what intermediate achievements should award points. A good way of yielding those points is by assigning one each time a high level skill is achieved. Every time a skill required for the current test has been correctly executed, a partial score is generated for that test. This is the way the RoboCup@Home organization solves the issue. For example, in the *Cocktail Party* RoboCup test a score is generated when the robot understands an order (human-robot interaction skill), finds and grasps a drink (robot cognition skill), delivers a drink to the correct person (robot cognition skill), etc...

d) 4 - Time it takes to solve the task.: Whether we like it or not, humans look to complete tasks as fast as possible. That is why quicker solutions should be scored higher that slower ones. At present, time scoring at competitions is implemented in a very simple way: just by allowing a certain amount of time to solve the task. Any solution within that time slot will be scored the same. Time execution is an issue to be included in a more complex scoring.

e) 5 - Setup time and setup work.: Neither setup time nor setup work is included in current scoring systems. At most, they provide a limited amount of time for setup prior to the test, that the team may use or not. Either the team uses it or not is only their business, but if a team does not require it, it is not rewarded. Setup work and setup time is something that matters to many users. However, it does not so much to others. Hence, even if it is agreed that requiring setup much be included in the scoring system, it is not agreed at all the amount of score that this must generate.

f) 6 - Price.: This is a point that has never being addressed in competitions, but will surely be in the future as the solutions mature. Of course lower price solutions should be preferred over more expensive ones.

g) 7 - Penalties.: It may happen that the system confuses information, people, objects, etc, and even if it performs the action, it does it completely wrong. Examples include the robot talking to furniture, or trying to grasp something ungraspable. Those factors evidence a total misunderstanding of the situation, for what some penalty factors should be included in the final score.

h) 8 - Importance of each factor.: Having addressed what we consider all the important criteria to take into account at a benchmarking for AAL systems, the final criteria is the weight that each of those factors has to have in the final score. This is a critical point, because if not correctly addressed it may produce unfair situations.

B. A fitness function for scoring AAL services

Creating the scoring function for an AAL service is like finding the fitness function of an evolutionary robotics algorithm [40]. Creating a fitness function for an evolutionary process, requires hard work of tuning it along the different experiments. That tuning is based on the experience of the researcher.

In evolutionary robotics, the researcher creates first a simple fitness function that accounts for all the factors he wants the solution to solve. He does not include in the function specific terms indicating a specific solution, just terms that indicate when the problem is solved. This allows the evolutionary algorithm find its own solution (this meets criteria 1).

Then he add terms to the function as long as he finds that they indicate requirements that have to be met. Usually those are specific requirements, or step stones that help to smooth the fitness function and avoid to make them very step (this meets criteria from 2 to 7). Very step functions prevent the evolutionary algorithm from finding a solution.

Finally, the researcher tunes the weights for each of those factors. The weighting of the factors is performed by the designer based on the results obtained in previous experiments and the experience of the designer (this would meet criteria 8).

The solution proposed here is similar to the evolutionary algorithm process. It consists in the creation of a complete score function based on the tasks the competition wants the robot to solve, either of higher or lower level (as described in section 2.1) and the achievement of the whole test. It has to include too a factor per each criteria of the previous section. Then, each of this factors has to be weighted in order to generate the final score. The next score function is proposed:

$$
\begin{aligned}
Score = & \Sigma \alpha_i low\_level\_skill_i \\
& + \Sigma \beta_j high\_level\_skill_j \\
& + \Sigma \gamma_k complete\_test_k \\
& + \lambda time \\
& + \delta setup \\
& + \mu price \\
& - \sigma \Sigma penalties_l
\end{aligned}
$$

The problem here is the selection of the weights

$$\alpha, \beta, \gamma, \lambda, \delta, \mu, \sigma.$$

It is not desirable to generate a subjective measure of the weights based on the own experience. This has demonstrated a bad approach at the RoboCup@Home competition (where the weights are completely arbitrary by the main board), being a source of many complains. Instead, we want to select the weights using an objective measure, as objective as possible. This would mean to systematize the whole competition into a set of equations and find the optimal weights. Since such kind of systematization is not possible, we propose a compromise, between being too objective (which cannot be performed in practice) and being too subjective (which depend only on our criteria): instead, we propose the use of a set of weights statistically generated (this is objective) from a set of polls asked to a considerable amount of people (which will provide subjective answers). The polls will ask people for their opinion on what should be more importantly scored on a test, how much they believe each factor should weight, etc[3].

IV. CONCLUSIONS

This paper has shown how the EvAAL competition can benefit from taking RoboCup@Home procedures and (robotic) systems. In the opposite direction EvAAL can also provide benefits to RoboCup@Home like for example the use of memory to remember things from previous executions (like remembering prices of things at the supermarket). Three potential benefits have been identified, and two different application scenarios have been proposed.

[3]The actual polls and results are at present work in progress

The paper also have addressed the problem of benchmarking for such competitions. There is no such thing as the perfect scoring system. In a general sense, the benchmarking solution proposed here tries to provide as less exploits as possible and eliminates subjectivity as much as possible.

Any scoring implementation that can be designed will have their own drawbacks, allowing participants to try to exploit them in order to have more points without actually solving the problem. This is also a common problem in evolutionary robotics where the evolving algorithms exploit the flaws of the whole system. But that is also what happens in nature. How does nature solve it? By taking a long time. As time passes more general solutions overtake the exploitation ones. A good exploit solution for today will not be valid anymore tomorrow because any other contestant will do the same at the next competition, requesting from participants to engineer more general solutions towards solving the actual problem in order to win.

REFERENCES

[1] J. K. Aggarwal and M. S. Ryoo. Human activity analysis: A review. *ACM Compututing Surveys*, 43(3):1–43, April 2011.

[2] G. Alenya and R. Tellez. The reem@iri 2012 robocup@home team description. In *Proc. CD of the 17th RoboCup International Symposium*, pages 1–8, Mexico, June 2012.

[3] J. A. Álvarez García, P. Barsocchi, S. Chessa, and D. Salvi. Evaluation of localization and activity recognition systems for ambient assisted living: The experience of the 2012 evaal competition. *Journal of Ambient Intelligence and Smart Environments*, 5(1):119–132, January 2013.

[4] A. Andrés, D. E. Pardo, M. Díaz, and C. Angulo. New instrumentation for human robot interaction assessment based on observational methods. *Journal of Ambient Intelligence and Smart Environments*, In Press, 2014.

[5] C. Angulo and R. Téllez. *Tendencias de la minería de datos en España*, chapter Distributed intelligence for smart home appliances, pages 1–12. Red Española de Minería de Datos, 2004.

[6] H. Arisumi, J.-R. Chardonnet, and K. Yokoi. Whole-body motion of a humanoid robot for passing through a door - opening a door by impulsive force -. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 428–434, St. Louis, MO, USA, Oct 2009.

[7] R. C. Arkin. Intelligent mobile robots in the workplace: Leaving the guide behind. In *IEA/AIE 1st International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems - Volume 1*, pages 553–561, Tullahoma, Tennessee, USA, 1988.

[8] P. Barsocchi, S. Chessa, F. Furfari, and F. Potorti. Evaluating ambient assisted living solutions: The localization competition. *IEEE Pervasive Computing*, 12(4):72–79, 2013.

[9] J. Bravo, R. Hervás, and M. Rodríguez, editors. *4th International Workshop on Ambient Assisted Living and Home Care, IWAAL*, volume 7657 of *Lecture Notes in Computer Science*, Vitoria-Gasteiz, Spain, December 2012. Springer.

[10] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.

[11] X. Chen, P. Stone, L. E. Sucar, and T. van der Zant, editors. *RoboCup 2012: Robot Soccer World Cup XVI*, volume 7500 of *Lecture Notes in Computer Science*. Springer, 2013.

[12] Y.-Y. Chen, J.-F. Wang, P.-C. Lin, P.-Y. Shih, H.-C. Tsai, and D.-Y. Kwan. Human-robot interaction based on cloud computing infrastructure for senior companion. In *IEEE Region 10 Conference, TENCON*, pages 1431–1434, Bali, Indonesia, Nov 2011.

[13] S. Chessa, F. Furfari, F. Potorti, J.-P. Lázaro, and D. Salvi. EvAAL: Evaluating aal systems through competitive benchmarking. *ERCIM News*, 2011(87), 2011.

[14] S. Chessa and S. Knauth. *Evaluating AAL Systems Through Competitive Benchmarking. Indoor Localization and Tracking*, volume 309 of *Communications in Computer and Information Science*. Springer, Berlin, Heidelberg, 2012.

[15] G. Doisy, A. Jevtic, and S. Bodiroza. Spatially unconstrained, gesture-based human-robot interaction. In *ACM/IEEE International Conference on Human-Robot Interaction HRI*, pages 117–118, Tokyo. Japan, March 2013.

[16] M. Donner, M. Himstedt, S. Hellbach, and H.-J. Boehme. Awakening history: Preparing a museum tour guide robot for augmenting exhibits. In *European Conference on Mobile Robots*, pages 337–342, Barcelona, Spain, May 2013.

[17] M. Eich, F. Grimminger, S. Bosse, D. Spenneberg, and F. Kirchner. Asguard: A hybrid legged wheel security and sar-robot using bio-inspired locomotion for rough terrain. In *IARP/EURON Workshop on Robotics for Risky Interventions and Enviromental Surveillance*, Benicassim, Spain, January 7-8 2008. Online-Proceedings.

[18] M. Eich, F. Grimminger, and F. Kirchner. Adaptive compliance control of a multi-legged stair-climbing robot based on proprioceptive data. *Industrial Robot: An International Journal*, 36(4):331–339, 2009.

[19] G. Fazekas, A. Toht, P. R. K. Zsiga, T. Pilis, and V. Dupourque. Cognitive-care robot for elderly assistance: preliminary results of tests with users in their homes. In *AAL Forum 2012*, Eindhoven, NL, 2012.

[20] D. Feil-Seifer, K. Skinner, and M. J. Mataric. Benchmarks for evaluating socially assistive robotics. *Interaction Studies*, 8(3):423–439, 2007.

[21] J. Forlizzi. How robotic products become social products: An ethno-graphic study of cleaning in the home. In *ACM/IEEE International Conference on Human-robot Interaction, HRI*, pages 129–136, Arlington, Virginia, USA, 2007.

[22] T. Frenken, M. Isken, N. Volkening, M. Brell, and A. Hein. Criteria for quality and safety while performing unobtrusive domestic mobility assessments using mobile service robots. In R. Wichert and B. Eberhardt, editors, *Ambient Assisted Living*, Advanced Technologies and Societal Change, pages 61–76. Springer Berlin Heidelberg, 2012.

[23] T. Fuxreiter, C. Mayer, S. Hanke, M. Gira, M. Sili, and J. Kropf. A modular platform for event recognition in smart homes. In *IEEE International Conference on e-Health Networking Applications and Services (Healthcom)*, pages 1–6, Lyon, France, June 2010.

[24] A. Gómez-Goiri, E. Castillejo, P. Orduña, X. Laiseca, D. L. de Ipiña, and S. Fínez. Easing the mobility of disabled people in supermarkets using a distributed solution. In J. Bravo, R. Hervás, and V. Villarreal, editors, *IWAAL*, volume 6693 of *Lecture Notes in Computer Science*, pages 41–48. Springer, 2011.

[25] P. Henry, C. Vollmer, B. Ferris, and D. Fox. Learning to navigate through crowded environments. In *IEEE International Conference on Robotics and Automation, ICRA*, pages 981–986, Anchorage, Alaska, USA, May 2010.

[26] M. Hoai and A. Zisserman. Talking heads: Detecting humans and recognizing their interactions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014. To appear.

[27] G. Hu, W.-P. Tay, and Y. Wen. Cloud robotics: architecture, challenges and applications. *IEEE Network*, 26(3):21–28, 2012.

[28] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. An affective guide robot in a shopping mall. In *ACM/IEEE International Conference on Human Robot Interaction, HRI*, pages 173–180, La Jolla, California, USA, 2009.

[29] R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, M. Nicolescu, and G. Bebis. Understanding human intentions via hidden markov models in autonomous mobile robots. In *ACM/IEEE International Conference on Human Robot Interaction, HRI*, pages 367–374, Amsterdam, The Netherlands, 2008.

[30] P. U. Lima, D. Nardi, L. Iocchi, G. Kraetzschmar, and M. Matteucci. RoCKIn@Home: Benchmarking domestic robots through competitions. In *International Conference on Advanced Robotics, ICAR*, Montevideo, Uruguay, November 2013.

[31] P. U. Lima and A. Winfield. 3rd workshop on robot competitions, challenges and benchmarking - exploring the synergies. https://sites.google.com/site/erf2014robocompworkshop/, 2014.

[32] C. Liu, C. Hu, Q. Liu, and J. K. Aggarwal. Video event description in scene context. *Neurocomputing*, 119:82–93, 2013.

[33] D. Lowet and F. van Heesch. Florence: A multipurpose robotic platform to support elderly at home. In *Workshop on Ambient Intelligence Infrastructures*, pages 21–24, november 2012.

[34] R. Madhavan, E. Tunstel, and E. Messina. *Performance Evaluation and Benchmarking of Intelligent Systems*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[35] I. Martí, J. P. Lázaro, and D. Salvi. EvAAL: Evaluating aal systems through competitive benchmarking. In *Actas de XIV Jornadas de ARCA. Sistemas Cualitativos y sus Aplicaciones en Diagnosis, Robótica e Inteligencia Ambiental*, pages 65–70, Salou-Tarragona, Spain.

[36] D. Martínez, G. Alenyà, P. Jiménez, C. Torras, J. Rossmann, N. Wantia, E. E. Arksoy, S. Haller, and J. Piater. Active learning of manipulation sequences. In *IEEE International Conference on Robotics and Automation, ICRA*, pages 5671–5678, Hong Kong, China, May 2014.

[37] S. McKeague, J. Liu, and G.-Z. Yang. Hand and body association in crowded environments for human-robot interaction. In *IEEE International Conference on Robotics and Automation, ICRA*, pages 2161–2168, Karlsruhe, Germany, May 2013.

[38] W. Meeussen, M. Wise, S. Glaser, S. Chitta, C. McGann, P. Mihelich, E. Marder-Eppstein, M. Muja, V. Eruhimov, T. Foote, J. Hsu, R. B. Rusu, B. Marthi, G. Bradski, K. Konolige, B. P. Gerkey, and E. Berger. Autonomous door opening and plugging in with a personal robot. In *IEEE International Conference on Robotics and Automation, ICRA*, pages 729–736, Anchorage, Alaska, USA, May 2010.

[39] H. Nakashima, H. Aghajan, and J. C. Augusto. *Handbook of Ambient Intelligence and Smart Environments*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[40] A. L. Nelson, G. J. Barlowb, and L. Doitsidis. Fitness functions in evolutionary robotics: A survey and analysis. *Robotics and Autonomous Systems*, 57(4):345–370, 2009.

[41] J. M. Ortman, V. A. Velkoff, and H. Hogan. *An Aging Nation: The Older Population in the United States*. United States Census Bureau, 2014.

[42] J.-Y. Puigbo, A. Pumarola, and R. Tellez. Controlling a general purpose service robot by means of a cognitive architecture. In *International Workshop on Artificial Intelligence and Cognition*.

[43] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.

[44] P. Rouanet, P.-Y. Oudeyer, F. Danieau, and D. Filliat. The Impact of Human-Robot Interfaces on the Learning of Visual Objects. *IEEE Transactions on Robotics*, 29(2):525–541, April 2013.

[45] S. Schneider, F. Hegger, A. Ahmad, I. Awaad, F. Amigoni, J. Berghofer, R. Bischoff, A. Bonarini, R. Dwiputra, G. Fontana, N. Hochgeschwender, L. Iocchi, G. Kraetzschmar, P. Lima, M. Matteucci, D. Nardi, and V. Schiaffonati. The RoCKIn@Home challenge. In *Proceedings of the 45th International Symposium on Robotics, ISR*, pages 192–199, Munich, Germany, 2014.

[46] S. Solaimani, H. Bouwman, and N. Baken. The smart home landscape: A qualitative meta-analysis. In *International Conference on Smart Homes and Health Telematics, ICOST*, pages 192–199, Montreal, Canada, June 2011.

[47] M. R. Tazari, R. Wichert, and T. Norgall. Towards a unified ambient assisted living and personal health environment. In *Ambient Assisted Living*, pages 141–155. Springer, 2011.

[48] M. Tenorth, K. Kamei, S. Satake, T. Miyashita, and N. Hagita. Building knowledge-enabled cloud robotics applications using the ubiquitous network robot platform. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 5716–5721, Tokyo Big Sight, Japan, November 3–7 2013.

[49] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 797–803, 2010.

[50] P. Trautman, J. Ma, R. M. Murray, and A. Krause. Robot navigation in dense human crowds: the case for cooperation. In *IEEE International Conference on Robotics and Automation, ICRA*, pages 2153–2160, Karlsruhe, Germany, May 2013.

[51] K. M. Tsui, H. A. Yanco, D. J. Feil-Seifer, and M. J. Mataric. Survey of domain-specific performance measures in assistive robotic technology. In *Workshop on Performance Metrics for Intelligent Systems, PerMIS*, pages 116–123, Gaithersburg, Maryland, 2008.

[52] J. van den Berg, S. Patil, J. Sewall, D. Manocha, and M. Lin. Interactive navigation of multiple agents in crowded environments. In *Symposium on Interactive 3D Graphics and Games, I3D*, pages 139–147, Redwood City, California, 2008.

[53] T. van der Zant and L. Iocchi. Robocup@Home: Adaptive benchmarking of robots bodies and minds. In *International Conference on Social Robotics, ICSR*, volume 7072 of *Lecture Notes in Computer Science*, pages 214–225. Springer, 2011.

[54] H. Vargas, E. Medina, D. Martínez, E. Olmedo, and G. Beristain. Human-robot interface using face detection and recognition, for the service robot, "donaxi". In *Intelligent Environments (Workshops)*,

volume 13 of *Ambient Intelligence and Smart Environments*, pages 242–251. IOS Press, 2012.

[55] M. Waibel, M. Beetz, J. Civera, R. d'Andrea, J. Elfring, D. Galvez-Lopez, K. Häussermann, R. Janssen, J. Montiel, A. Perzylo, B. Schiessle, M. Tenorth, O. Zweigle, and M. R. V. de Molengraft. Roboearth: a world wide web for robots. *Robotics Automation Magazine, IEEE*, 18(2):69–82, 2011.

[56] Z. Wang, K. Mülling, M. P. Deisenroth, H. Ben Amor, D. Vogt, B. Schölkopf, and J. Peters. Probabilistic movement modeling for intention inference in human-robot interaction. *International Journal of Robotics Research*, 32(7):841–858, 2013.

[57] T. Wisspeintner, T. van der Zan, L. Iocchi, and S. Schiffer. Robocup 2009. chapter RoboCup@Home: Results in Benchmarking Domestic Service Robots, pages 390–401. Springer-Verlag, Berlin, Heidelberg, 2010.

[58] T. Yoshida, K. Nakadai, and H. G. Okuno. Two-layered audio-visual speech recognition for robots in noisy environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 988–993. IEEE, 2010.

[59] C. L. Zhu, R. Boldu, C. de Saint Germain, S. X. Ubach, J. Albo, and S. Pfeiffer. The reem@iri 2014 robocup@home team description. In *Proc. CD of the 18th RoboCup International Symposium*, pages 1–8, Eindhoven, June 2013.