Towards Supervised Autonomy in Cloud Social Robotics

* Peter SINČÁK, Gergely MAGYAR, Center for Intelligent Technologies, TU Košice, Slovakia
Ivana ČERNAJOVÁ Senior House of Juraja Schoppera, Slovakia
Filippo CAVALLO Scuola Superiore Sant'Anna, Italy
Kaori YOSHIDA, Department of Human Intelligence Systems, Kyutech, Japan

1. Introduction to Cloud Social Robotics

Considering the developments in robotics, the definition of the notion "robot" can change over time. Information technologies (hereinafter IT), more particularly cloud computing, had a great impact on social robotics. Socially assistive robotics (hereinafter SAR) has many aspects, but the major one is human-robot interaction (hereinafter HRI), where the 'H' stands for user studies that means the human perspective; the 'R' represents the technological and other characteristics of the robot. Gaining insights into the 'I' (the nature of interaction between humans and robots) is possible only by investigating both 'H' and 'R'. Considering this, we can formulate the key challenge of HRI as follows [1]: "HRI is the science of studying people's behavior and attitudes towards robots in relationship to the physical, technological and interactive features of the robots, with the goal to develop robots that facilitate the emergence of human-robot interaction that are at the same time efficient (according to the original requirements of their envisaged area of use), but are also acceptable to people, and meet the social and emotional needs of their individual users as well as respecting human values".

HRI is a relatively young discipline evoked by the increase in the availability of robots and people's exposure to them. Human-robot interaction has become a part of everyday life through robotic toys and household appliances (robot vacuum cleaners or lawn movers). Robots found their applications in real-world application areas, such as rehabilitation, eldercare, or robot assisted therapy.

Socially assistive robotics has three major multidisciplinary parts as it can be seen in Figure 1. This paper focuses on the IT and artificial intelligence (hereinafter AI) part of this problem. It is believed that the significant progress in IT and AI could be used in HRI to generate new approaches for the assistive domain of social robotics.

The major challenge related to creating new business models in social robotics is connected to the question what a robot is. In the concept of 'Reality 2.0' from Japan and 'Industry 4.0' used mainly in Europe we can consider the possibility of connecting cyber and physical systems in social robots.

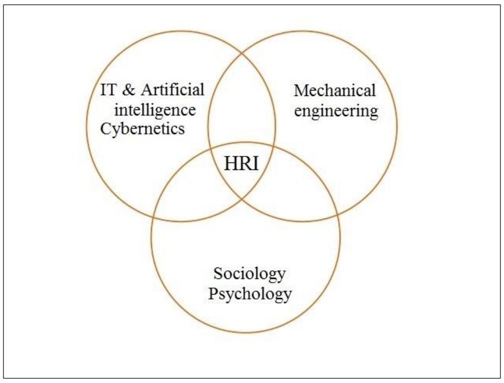

Fig. 1 Multidisciplinarity of HRI

Understanding these approaches, we can reformulate the definition of a robot and accept the fact that a connection between a virtual and a real robot can be really fast.

This assumption is supported by the technological progress in Wi-Fi networks towards 5G and cloud computing. Based on this, the notion 'robot' can have the following meanings:

a) **Classical** understanding of a robot
b) **Virtual robot** (hereinafter VR) on the cloud (or server) which is constantly connected to one or many real robots (hereinafter RR). This means that VRs with different parameters can control many RRs, which can lead to personalization in HRI.
c) **Omni-present** (ubiquitous robot concept) approach when a VR or a RR can be connected to an intelligent space including external cameras, social networks, or knowledge bases. Data fusion and knowledge discovery from big data are the key elements of this approach.

Based on the considerations above we can change the concept of HRI in assistive robotics to HRE-I (Human-robot-environment interaction), where the environment is the source of big data which helps to personalize the robot's services. In Figure 2 one can see the basic concept of a new way of interaction in SAR. The key element is the soft agent which is considered as a VR connected to a RR (in this case a Pepper). The main challenge in such an architecture is to make the VR semi- or fully-autonomous.

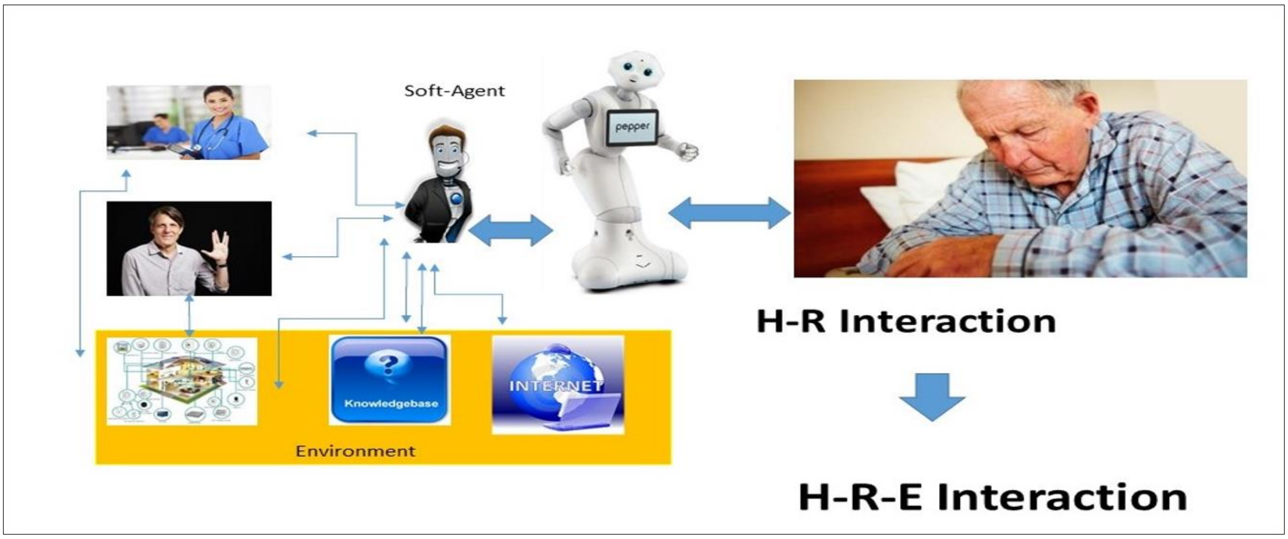

Fig. 2 Basic concept of HRE-I for socially assistive robotics

2. From Wizard of Oz to Autonomous VR

Machine (or robot) autonomy is a complex issue and an even complicated notion. In our case we assume that the VR is teleoperated by a human. This approach in the HRI community is called the Wizard of Oz method. The basic approach for measuring the autonomy of a robot is to consider task-based autonomy. If we consider that a task is carried out by a human in collaboration with a machine, we can state that:

$$GIQ_{TASK} = HIQ_{TASK} + MIQ_{TASK} \qquad (1)$$

where GIQ_{TASK} is a task-dependent global intelligence quotient, always set to 1; HIQ_{TASK} is a task-dependent human intelligent quotient, from the interval <0, 1>; MIQ_{TASK} is a task-dependent machine intelligence quotient, from the interval <0, 1>,

According to the above-mentioned, when HIQ_{TASK} is 1 and MIQ_{TASK} is 0, the VR is entirely controlled by the human teleoperator. When MIQ_{TASK} is 1 and HIQ_{TASK} is 0 then the VR is fully autonomous and the human teleoperator is just supervising the VR.

The main scientific challenge is how to create an autonomous VR personalized to the human's needs. Albeit there are several approaches to do so, we have been investigating various methods of reinforcement learning for learning from teleoperation in the cloud environment.

One of the most popular reinforcement learning (hereinafter RL) algorithms is Q-learning developed by Watkins in 1989 [2]. This method is an off-policy TD (temporal difference) control algorithm which calculates the quality of a state-action combination ($Q: S \times A \rightarrow R$). It is based on the theory of Markov Decision Processes (MDP), which states that any state s_{t+1} occupied by an agent is a function only of its last state and action: $s_{t+1} = f(s_t, a_t)$, where $s_t \in S$ and $a_t \in A$ are the state and action at time step t [3]. The model of the problem includes an agent, states (s) and actions (a) which are connected to states. By performing an action, the agent changes its state which provides it with a reward (r). The goal of the agent is to maximize the total reward by learning which actions to perform in each state. In general, the action which is optimal for the given state has the highest long-term reward. The reward is the weighted sum of the expected values of rewards of future steps starting from the current state. Before learning, the first value of Q is given by the developer (usually 0), later on the values are stored in a table. The Q-values are calculated by the following formula [4]:

$$\begin{aligned} Q_{t+1}(s_t, a_t) = \;& Q_t(s_t, a_t) + \alpha(s_t, a_t)[r_{t+1} \\ & + \gamma max Q(s_{t+1}, a) - Q_t(s_t, a_t)] \end{aligned} \qquad (2)$$

where r_{t+1} is the reward after performing action a_t in state s_t. $\alpha(s_t, a_t)$, $(0 < \alpha \leq 1)$ is the learning rate which determines whether a new information will override the existing one. The learning rate can be different for various states. Γ is the discount factor which determines the importance of future rewards [4].

We have tried different types of RL and ported them to a cloud computing environment using the Microsoft Azure platform. This kind of implementation has the following advantages:

- **Software as a Service (SaaS)** – is a capability provided to the user to use an application which is running on a cloud infrastructure. By infrastructure we mean a collection of software and hardware which enables the five characteristics of the cloud model (on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service). The application is available from a various range of devices through a thin client or a program interface. The developer doesn't manage or control the cloud infrastructure, such as network, servers, operating system, etc.

- **Scalability** – is the application's ability to continue to function without any errors when it is change in size to meet the users' need. In some cases it is done automatically, but there is an option to define the rules controlling this procedure manually by the developers of the application. From the user's point of view these capabilities can be seen as unlimited and be appropriated at any time.
- **Accessibility** – the results of the learning process (in our case the trained Q-tables) are always available and can be used in similar scenarios with the given subject.
- **Personalization** – the trained Q-tables for each individual can be used to search for patterns, which can serve as a basis for creating a personalized robotic behavior.
- **Speedup in the learning process** – since the cloud infrastructure optimizes the deployed algorithms the learning process can be sped up significantly.

We think that the above-mentioned characteristics can bring a new dimension to the use of reinforcement learning in or out of social human-robot interaction.

3. Cognitive Stimulation Therapy with Elderly

To prove the usability of our approach for learning social action selection from the Wizard in a real-world social HRI scenario we used a robot as a coach for cognitive stimulation therapy in an elderly care facility. In a similar work [5] a robot was used for cognitive stimulation therapy for individuals suffering from Mild Cognitive Impairment (MCI) and/or Alzheimer's disease. The goal of the interactions was to increase the participants' cognitive attention by playing different levels of difficulty. The games played were tailored to the needs of each individual to address their different cognitive disabilities. There were two experiments where the feasibility of the system was tested. The first one consisted of an interaction between the robot and the user with the goal of helping the subject to improve or maintain their cognitive attention through encouragements in a music-based stimulation game. In the second one the robot was present during the game and suggested a new scenario for the individual primarily focused on the problems revealed by the current game.

3.1. Hypotheses of the Experiments

In our experiments we wanted to test the following hypotheses:

H1: Our modified interactive reinforcement learning algorithm is able to increase the robot's level of autonomy in a real-world social human-robot interaction scenario.

H2: The elderly will be able to accept and build a positive relationship with the robot even only after two interactions with it.

3.2. The methodology

In our setting the participants played two games called "Guess the fruit" and "Guess the animal" with a NAO robot. In the games the robot chose a letter and asked the participant to say a fruit/animal which name starts with the same letter, e.g. for A – apple/antelope. When the answer was correct, the robot celebrated, otherwise it gave a hint about a possible correct answer. The game lasted until the subject gave correct answers for all letters.

The technical description has to be divided into two parts, since in the experiments the robot was controlled remotely using the Wizard of Oz technique – however, our learning algorithm was running in the background to increase the robot's level of autonomy.

The session itself took place in a room, where the subject, his/her caregiver and the robot were present. The interaction was monitored by three cameras. The technical setup of the above can be seen in Figure 3.

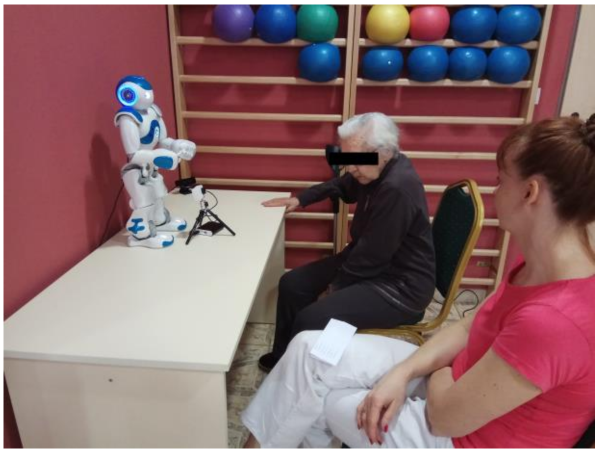

Fig. 3 The experimental setup

The Wizard in the other room had a simple graphical user interface connected to the cloud at his disposal for controlling the robot and the learning process. To test our learning algorithm in a real-world social human-robot interaction scenario we defined 4 states of the subject and 6 types of actions of the robot, while the reward for each action ranged between 0 and 100. The 4 states were the following:

- **Correct answer** – the state when the participant answered correctly
- **Incorrect answer** – the state when the participant answered incorrectly
- **Doesn't know** – the state when the participant didn't know the answer
- **Other** – defined a state other from the above-mentioned, e.g. the state when the robot finished its applause

During the cognitive experiments 6 types of actions were the following:

- **Easy question** – the action for asking an easy question. As an easy question we defined the letters for which it is simple to find a fruit or an animal. When determining the difficulty of each question we analyzed previous sessions where the games were played with a caregiver.
- **Question of medium difficulty** – the action for asking a medium difficulty question. As a medium difficulty question we defined the letters for which it was harder to find a fruit or an animal, but there were more than on possibilities.
- **Hard question** – the action for asking a hard question. As a hard question we defined letters for which it was hard to find a fruit or an animal. In most cases there was just one possible correct answer.
- **Help** – for each letter we created a short description of a possible solution.
- **Applause** – the action executed after each correct answer. We created 6 different applauses for both games.
- **Sorrow** – the action executed after each incorrect answer. We created 6 different expressions of sorrow for both games.

4. Results

The experiments were carried out in an elderly care facility, where the participants live and used to play cognitive exercise games with their caregivers twice a week.

During our stay at the facility the elderly played two games with the robot. The first game was played by 12 participants (3M/9F, average age 78.5, 56 to 94), and the second by 10, since two of the participants had to visit their doctors. We also have to note, that the participants were mentally healthy, except for one subject who shows the signs of early dementia, however, the caregivers found her suitable for our experiments.

The results from the robot's autonomy view describe the efficiency of the learning algorithm, comparing the percentage of actions executed autonomously (these actions were not chosen manually, but only approved by the Wizard) in the first and second session.

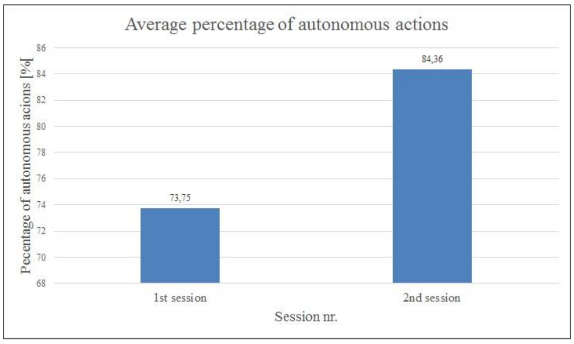

Fig. 4 Averaged level of autonomy of VR on the Azure Cloud.

The results shown in Figure 4 summarize the outcomes of the 2 interactions with 10 participants, who were able to attend both sessions (2 males /8 females, average age 78.5, in interval from 68 to 94). The results clearly show the increase of autonomy of the VR and decrease the of the Wizard's cognitive load. That means, that we increase the MIQ$_{TASK}$ and decreased the HIQ$_{TASK}$.

The developed cloud social robotics framework provides an ability to collaborate and use it with different users and robots (currently supports NAO, Pepper, DORO, Milo and Q.bo). In the near future we plan to include the Telenoid in collaboration with ATR in Japan.

5. Acknowledgments

This research work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-015-0731 and research project is supported from 07-2016 to 06-2019 also with Research supported by the National Research and Development Project Grant 1/0773/16 2016 – 2019 "Cloud Based Artificial Intelligence for Intelligent Robotics". Finally, the research is also supported by Microsoft Research in Cambridge, UK with Azure Cloud services for 2017-2018.

6. REFERENCES

[1] Kerstin Dautenhahn: "Human-robot interaction,"

[2] Christopher Watkins and Peter Dayan: "Q-learning," Machine learning 8, no. 3-4, pp. 279-292, 1992.

[3] Christopher Watkins: "Learning from delayed rewards," PhD thesis, University of Cambridge, 1989.

[4] Richard S. Sutton and Andrew G. Barto: "Reinforcement learning: An introduction," Vol. 1, no. 1. Cambridge: MIT press, 1998.

[5] Adriana Tapus, and Alexandra-Maria Vieru: "Robot cognitive stimulation for the elderly," International Work-Conference on the Interplay Between Natural and Artificial Computation, pp. 94-102. Springer, Berlin, Heidelberg, 2013.