A Cognitive Architecture for Human-Robot Teaming Interaction

Antonio Chella[1,2], Francesco Lanza[1], and Valeria Seidita[1,2]

[1] Dipartimento dell'Innovazione Industriale e Digitale
Universitá degli Studi di Palermo
[2] C.N.R., Istituto di Calcolo e Reti ad Alte Prestazioni, Palermo, Italy
name.surname@unipa.it

Abstract. Human-robot interaction finalized to cooperation and teamwork is a demanding research task, both under the development and the implementation point of view. In this context, cognitive architectures are a useful means for representing the cognitive perception-action cycle leading the decision-making process. In this paper, we present ongoing work on a cognitive architecture whose modules consider the possibility to represent the decision-making process starting from the observation of the environment and also of the inner world, populated by trust attitudes, emotions, capabilities and so on, and the world of the other in the environment.

Keywords: Human-robot interaction · Cognitive Architectures.

1 Introduction

A human-robot interaction system is created for different purposes, from assisting in home, healthcare or learning to safety or search and rescue scenarios and so on; in any case, it is a complex system to design and develop. If also, we consider, a human-robot team where the robot and the human have to autonomously cooperate, communicate and collaborate to reach a common and shared objective, we are talking of a more complex system. Here the term "complex" is used to mean that, in such a context, the robot, as well as the whole system, does not have a behavior that can be analyzed and implemented as a sum of the parts because it assumes an emergence of behaviors at runtime. For instance, let us suppose to have a team made up of a human and a robot who has to carry out a task, known to both of them, interacting with each other. During the design phase, all the actions that each of the two must perform and all the communications that they must exchange during the interaction can be set. During the execution phase, if the task is to be carried out in a dynamic environment, the interactions between the robot and the human and between them and the environment inevitably change the state of the world that can thus provide new constraints or requirements for achieving the initial goal. The actual behavior of the whole system comes out at runtime and the team members need to be able to respond to changes efficiently.

The most significant difficulties in these types of systems are related to equipping the robot with the ability to select at runtime the best action to perform to achieve the team's goal. These types of problems are often studied and implemented by looking at their human counterpart, in our case the human-human team. The question is: how does a human being act in a team whose goals he knows and shares in a changing environment? And, how can this be reported in a human-robot team? Usually, a human being grounds his decisional process on a set of factors, such as knowledge of the surrounding environment and knowledge of a set of possible plans allowing to achieve a goal.
These factors, which could be described as reactive, occur when the situation being addressed is known and the changes are foreseeable before an action plan is established. When one is engaged in a cooperative and dynamic context some psychological factors intervene, that is, the whole set of internal states that trigger a decision and are closely related to the mental state of the human being, the knowledge of himself and the other elements of the team. For example, one element of a team may understand to be not able to perform a certain action that can lead to the achievement of the common goal. It cannot, also, want to do an action, and delegate it to the other team member based on a certain level of trust in the skills of the other. Or it could be in such an emotional state (for example euphoria) to want to do more than what was told during the design phase and so propose to the other component to carry out some actions or even decide to act independently on the behalf of the other.

In this context, the human behaves, in all respects, as a cognitive agent so a research field that is spreading nowadays is that of cognitive architectures. In recent years, various cognitive architectures have been studied and have served to examine how human beings behave, but they have also been adopted as the basis for the implementation of robotic systems that act like humans.

This paper presents an ongoing study focused on the use of self-modeling and the theory of mind in order to propose a cognitive architecture. It would allow modeling a human-robot teaming system that cooperates to achieve a common goal and, as well as a human team, applies a decision-making process that is driven from the objective situation of the environment, but also from the knowledge that each cognitive agent has of itself and the other members of the team.

In the rest of the paper, we set state of the art in the cognitive architecture of human-robot interaction systems and the motivation of our work, then we present our proposal for a cognitive architecture and finally we draw some discussions and conclusions.

2 Cognitive Architectures for HRI

Modeling and understanding cognitive processes are the primary aim of the research on cognitive architectures. A cognitive architecture determines the structure of a computational cognitive model, applied in a generic domain, by underlying the infrastructure of an intelligent system.

Typically, cognitive architectures include modules representing the short and long-term memory and the functional processes operating on memory modules for realizing learning and action mechanisms. Several cognitive architectures have been proposed in the literature; our work starts from the analysis of a lot of them in order to deduce the elements suitable for our application context and whether and how they needed to be extended and integrated. We have been driven by the assumption that cognitive architectures model intelligent systems where components' behaviors are not established and coded at design time but arise from perception and knowledge.

Among the most known cognitive architectures, we may count those inspired by psychology and biology (for instance CLARION, SOAR, ACT-R ...) and those by agent technology, like for instance LIDA [2, 6, 10, 13]. The fundamental principles they underpin are that every autonomous agent, be it human, animal, robot, software component/agent etc., has to continuously sense the environment and reason on it in order to be able to select the appropriate action to reach an objective (action-perception cycle [7]). The previous scenario is the simplest one: an agent that has to be endowed with the capability of answering in a specific context exhibiting the right action. Thus, regardless of the context, a cognitive cycle can be schematically outlined as the module receiving all the sensorial data and the one processing them thus resulting in the corresponding action.

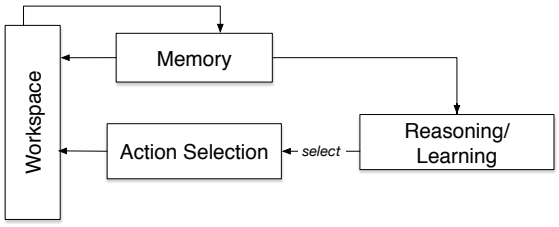

Fig. 1. Common modules in cognitive architectures.

In addition to these two modules, there is a basic module that manages everything related to data storage for further processing. This allows to represent, and therefore to implement, a system in which the decisional process passes from the elaboration of a set of data, relative to everything that is perceived, memorized in the system. In existing architectures, the memory is mainly divided into short-term memory and long-term memory. The short-term memory stores all the facts and elements relating to the current moment. The long-term memory contains all data and events that are held indefinitely. Long-term memory is further divided into explicit and implicit memory in all its different forms (episodic, semantic, autobiographical...) already widely debated in the literature [11, 1].

So, the memory module plays a fundamental role in reasoning on and learning the current situation in order to trigger the decision process.

Nevertheless, frequently human-robot teaming applications are not so simple to be sketched through these three simple modules but it is necessary, at least,

that the memory mechanism also refers to all the evidence coming from perception and from knowledge of self and from the perception and knowledge of all the other agents present in the environment. This means that, in a human-robot teaming context, architectures must contain the modules useful for the representation of itself, of the (physical) world around and of the other, including all the mental states that this fact entails.

To date, most architectures base their decision and learning process only on the concept of stored data or fact and not on the notion of mental state, and also the implementation part of most known cognitive architectures is still missing. Our contribution lies in the creation of a cognitive architecture in which memory also contains all the information about the mental state so that the perceive-act cycle becomes what we call the perceive-proact cycle. Besides, the architecture we propose can be easily mapped to a BDI agent architecture [17, 8] for the effective implementation of a cognitive agents system.

In the next paragraph, we explain how the architecture we propose includes the modules to model and represent a human-robot teaming system.

3 Modeling and Representing Autonomous and Adaptive Interactions

In Fig. 2 we illustrate the cognitive architecture we propose for realizing human-robot teaming interaction. The architecture is composed of the several known modules from general cognitive architecture and is enriched with the modules devoted to employing decision functionalities.

Our starting point is the generalization of all the characteristics of the cognitive architectures studied that led, together with the in-depth analysis of LIDA, to the representation given in Fig. 1. The figure shows what has been said above regarding a perception-action cycle; the workspace is the module used for interaction with the work environment.
In LIDA the workspace module takes into account two different facts: the changing environment, in the cognitive cycle the agents are able to make an internal representation of the world; all this helps agents in selecting the actions to be done also taking into account the conscious part of the agent. Consciousness in LIDA is the mental counterpart of attention, that is, of the perception process. LIDA gives a complete representation of a cognitive cycle but it is not suitable to represent interactions in teaming because it lacks the part related to the representation of the self and to the representation of the other.

Before going into the details of the proposed architecture, it is worth to make a hint to some necessary starting hypotheses. As was said in the introduction, a human-robot system working in team to achieve a common objective is a complex system. A complex system, by definition, is a system made up of complex components that interact with each other and with the surrounding environment. The global behavior and characteristics of the system emerge from interactions. The behavior of a complex system cannot be analyzed and implemented as the sum of the single components, the system must be seen as a whole. In our case,

we consider the system as composed of a set of cognitive agents, living in an environment made by inanimate objects and cognitive agents at the same time. Analysis and design of such a system must begin by choosing the perspective from which we look at the system.

For example, if the perspective is that of the cognitive agent (the robot), it will see the system as the set of inanimate objects on which he can act. It will see itself with everything he has inside (his goals, his capabilities, his mental states) and he will see the other (any other cognitive agent) with everything he has inside and with the set of objects on which the other can act. So the man-robot system is, to all intents and purposes, a system of systems in which the environment is not external like something with which he only interacts but is an integral part of the system itself.

This is the main difference between our architecture and the existing ones. Such a feature allows us to represent and then implement a decision-making process based on a series of factors similar to human ones, such as the sense of self, elements of the theory of the mind, trust, emotions and everything can generate a mental state. In so doing we are able to realize a change of design paradigm, from design time to runtime, already studied and analyzed in some previous works [12, 5], and also allows us to create a system that can adapt to runtime new situations.

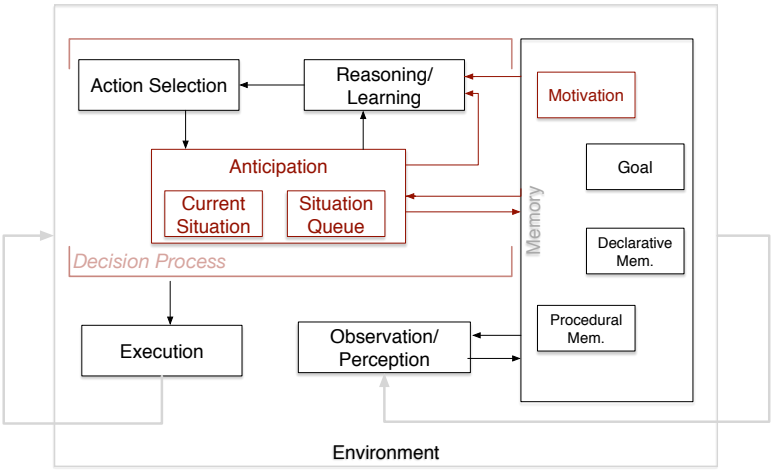

Fig. 2. The Proposed Cognitive Architecture for Human-Robot Teaming

Fig. 2 our architecture is shown. As it can be noted, the cycle Monitor, Analyzing, Planning and Executing (MAPE), which is the basis of all the implementations of complex autonomous systems [3], is realized within the four main modules (the ones in black) along with the memory module.

In [15, 14], the authors define a cognitive system as, generally, an agent able to perceive the environment and learn from experience. At the same time, a cognitive agent is able to anticipate the outcome of its action and to adapt to changing situation in the environment by acting driven by its own motivation (so called pro-activeness in the field of multi-agent systems [9, 16, 4]). These latter two parts of the definition imply specific architectural modules when we study robots that interact with a human in a collaborative fashion. Indeed, in the architecture we propose, the decision process is centered in the ANTICIPATION and MOTIVATION modules. Moreover, we explicit the representation of GOAL; it is part of the memory of the overall system and embodies the state of the world the robot-human team wants to achieve. A goal is in and tightly linked to the memory because we claim that, for a cognitive agent be able to act autonomously, it has to be configured along with all the elements of the knowledge useful for successfully pursuing it. We mean among the others: agent's skills, knowledge on the environment, pre-condition for the commitment, knowledge on the possible actions to perform for pursuing it, all the possible factors contributing or preventing it. These factors may be already stored in the memory or have to be acquired and stored at runtime.

The DECISION PROCESS part of the architecture is centered on ANTICIPATION and MOTIVATION. According to our architecture, the robot does not act only after the reasoning process based on a certain data stored in the memory, but also and mainly after evaluating the anticipation of its actions. During the anticipation process, the current situation is generated; it represents the state of the world corresponding to the currently selected action. The current situation is elaborated on the basis of motivations, goals and all those elements are in the memory thus getting the execution launched. A queue of possible situations is also created, intended as a set of pre-conditions, objectives and knowledge to achieve them, and postconditions on the objectives. The robot can draw on all these elements at any time to respond to changes and still maintain its initial target. The MOTIVATION module is the one triggering the anticipation and the action selection. It is the core of the decision process, here all the information and process for elaborating mental states reside and it is the module devoted to the representation of inner and word of each cognitive agent. Through this module, therefore, it is possible to make decisions about the actions being conveyed by the sense of self, by the ability to attribute mental states (belief, desire, intention, knowledge, capabilities) to oneself and to others and by the understanding that others have different mental states, by emotions, by the level of trust in the abilities (or more generally by trust) of others and of oneself.

The experiments [3] we are performing in our laboratory are confirming that with these modules, the architecture we propose can be easily mapped into a BDI agent architecture thus giving the possibility to implement a human-robot interaction system with a multi-agent system made of cognitive agents.

[3] Showing the experiments is not in the scope of this paper.

4 Discussions and Conclusions

Designing and implementing a human-robot interaction system in a dynamic environment and for tasks that require team organization is a challenging issue. The literature provides a rich set of cognitive architectures that allow modeling the human-robot interactions through the classical perception-action cycle of a cognitive agent. However, the application domain we are working on is such that considering and using the action-perception cycle only is not enough and also, most existing cognitive architectures do not provide a usefully implemented counterpart. In this work, we have therefore generalized and extended the existing architectures in order to obtain the modules that meet our requirements, namely to create a human-robot interaction on the basis of the human-human interaction in a team that cooperates in the same highly evolving environment to achieve a known and shared goal.

The main contribution of this work is the module for the anticipation of the situation and the one for the representation of the internal state of the robot and for the representation of knowledge about the internal state of the other. In this way, we can integrate the use of self-modeling and the theory of mind in team interaction and besides, we can scale the use of this module including a whole set of mental states that are typical in the human, therefore emotions, levels of trust in oneself and the others, etc. It was also fundamental to make a change in the abstract representation of the environment, from the element with which the robot interacts, as a portion of the world outside, to the element that is the world together with the robot itself as in a sort of recursive structure.

In this way, for example, the work of the execution module on the environment and therefore on any elements of the environment can be configured, for example, in the delegation an action from the robot to the human being when it cannot do an action for any reason generated by the knowledge of himself and the environment.

Finally, the module we have called MOTIVATION, allows us to verticalize the architecture from a theoretical level to a system level with a mapping one by one with a BDI agent architecture. We are already doing several experiments on this and in the future, we will realize the whole system in which the internal state of the robot completely conveys decisions on actions. We expect, for example, that a robot at the beginning of its activities, which has learned few elements of the surrounding world and of the human being with whom it is interacting, has little margin for autonomous choices on the actions to be taken and is only able to implement those given at design time. After a certain number of interactions, it will have acquired knowledge, a certain level of trust in the other and be able to act autonomously.

Acknowledgment. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-17-1-0232.

References

1. Anderson, J.R.: The architecture of cognition. Psychology Press (2013)
2. Anderson, J.R., Matessa, M., Lebiere, C.: Act-r: A theory of higher level cognition and its relation to visual attention. Human-Computer Interaction **12**(4), 439–462 (1997)
3. Andersson, J., Baresi, L., Bencomo, N., de Lemos, R., Gorla, A., Inverardi, P., Vogel, T.: Software engineering processes for self-adaptive systems. In: Software Engineering for Self-Adaptive Systems II, pp. 51–75. Springer (2013)
4. Bordini, R.H., Hübner, J.F., Wooldridge, M.: Programming multi-agent systems in AgentSpeak using Jason, vol. 8. John Wiley & Sons (2007)
5. Cossentino, M., Sabatucci, L., Seidita, V.: Towards an approach for engineering complex systems: Agents and agility. In: Proceedings of the 18th Workshop "From Objects to Agents" Scilla (RC), Italy, June 15-16, 2017. (2017)
6. Franklin, S., Madl, T., D'Mello, S., Snaider, J.: Lida: A systems-level architecture for cognition, emotion, and learning. IEEE Transactions on Autonomous Mental Development **6**(1), 19–41 (2014)
7. Fuster, J.M.: Upper processing stages of the perception–action cycle. Trends in cognitive sciences **8**(4), 143–145 (2004)
8. Georgeff, M., Rao, A.: Rational software agents: from theory to practice. In: Agent technology, pp. 139–160. Springer (1998)
9. Jennings, N.R., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. Autonomous agents and multi-agent systems **1**(1), 7–38 (1998)
10. Laird, J.E., Newell, A., Rosenbloom, P.S.: Soar: An architecture for general intelligence. Artificial intelligence **33**(1), 1–64 (1987)
11. Newell, A., Simon, H.A., et al.: Human problem solving, vol. 104. Prentice-Hall Englewood Cliffs, NJ (1972)
12. Seidita, V., Cossentino, M.: From modeling to implementing the perception loop in self-conscious systems. International Journal of Machine Consciousness **2**(02), 289–306 (2010)
13. Sun, R.: The importance of cognitive architectures: An analysis based on clarion. Journal of Experimental & Theoretical Artificial Intelligence **19**(2), 159–193 (2007)
14. Vernon, D.: Artificial cognitive systems: A primer. MIT Press (2014)
15. Vernon, D., Metta, G., Sandini, G.: A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. IEEE transactions on evolutionary computation **11**(2), 151–180 (2007)
16. Wooldridge, M.: An introduction to multiagent systems. John Wiley & Sons (2009)
17. Wooldridge, M., Jennings, N.R.: Intelligent agents: Theory and practice. The knowledge engineering review **10**(2), 115–152 (1995)