Agents' behavior and its conceptual framework

Stefano Borgo[orcid.org/0000−0001−6001−2765]

Laboratory for Applied Ontology, ISTC CNR, Trento (IT)
stefano.borgo@cnr.it
https://www.loa.istc.cnr.it

This note concentrates on the notion of agents' behavior, as understood today, and suggests expanding it to study future socio-technical systems. The proposal is to take an ontological viewpoint to discuss anew topics like: how to understand behavior, how to connect behavior to an agents' type, how to structure a broader framework for behavior studies. Due to space limitations, the presentation focuses on motivations and gives only a few guiding ideas. Overall, the aim is to foster the transformation of today's socio-technical systems in inclusive systems where humans and artificial agents cohabit and, possibly, develop social practices, norms and other forms of collaboration [1].

Several foundational ontologies can be used today to formalize so-called *conceptualizations* of reality [2]. These ontologies adhere to distinct principles and categories, leading to mutually inconsistent formalizations. This is unavoidable as there are several coherent and philosophically sound ways to explaining reality. While it is important to have sound and clearly motivated systems, their philosophical antagonism may hinders progress without being of much interest to science. To avoid this problem, here ontologies are seen as mere conceptual systems disconnected from philosophical views and related quarrels. Also, since humans understand and use most, if not all, these ontological systems (one may find bizarre and aberrant some ontology), one should be free to use her preferred ontological approach. For this reason, I advocate the construction of a general high-level module for behavior not specifically tuned to an ontology.

A problem towards inclusive socio-technical systems

The purported view of future socio-technical systems shows humans and artificial agents forming integrated communities with shared social practices, norms and conventions [3]. The idea that our society will evolve to accomodate cyborgs and robots is today taken for granted, and indeed this process has started long ago in fields like industry, transportation services, medicine and domotics. However, the potentialities of autonomous robots, whose capabilities are quickly developing, suggest considering them not as devices but as another species [4]. The direct implication is that robots can, and likely will, develop independently of humans' interests and control. Progress in this direction is already documented in the domain of evolutionary robotics and learning [5, 6], and includes the creation of human-independent languages [7].

To take the most out of this evolution, humans can anticipate how to interact with the new species [1]. One building bock to foster a positive evolution, especially if a shared language is lacking, is the development of a rich conceptual framework for behavior description and understanding. I am not advocating studies of generic behavior as in physics. Rather, I am looking at environment- and interaction-driven behavior where at least one involved entity is an agent.

Humans have already experience in the understanding of animal behavior through observational data interpreted on the bases of, e.g., developmental psychology, cognitive science, behavioral ecology, ethology, neuroscience. This research can be expanded to behavioral studies in robotics and artificial intelligence. To start this, one needs to integrate the engineering viewpoint (aiming to generate, and possibly control, robots' behavior), and the human-robot interaction viewpoint (aiming to develop collaborative and adaptable behavior).

Toward an ontology-bases behavior framework

Any system of applied ontology, by which we mean a consistent set of concepts for understanding and describing reality, is based on ontological analysis: a process to identify the relevant entities and properties in the domain at stake. Distinct ontologies differ on the concepts they use but ontological analysis remains key to identifying types of behavior and their characteristics.

The core notion, behavior, is here taken to be the mutual relationship between an entity and its environment. This means that one can likewise talk of the behavior of a person, a computer, a tree or a stone, and that the behavior depends on the entity and the entity's environment (at the time of interest), and thus is a relational quality. Briefly, the behavior manifests the way the entity affects the environment and the way the entity is affected by the environment. For the purposes of this note, I constrain the attention to embodied agents. More precisely, I define *agent behavior* to be a relational quality that inheres in a physical agent and whose manifestation is relative to the agent's environment. By *environment* I mean a situation, that is, a spatio-temporal region and every object and process in it. By *agent's environment* I refer to an environment that contains the agent and is restricted to all the objects with which the agent could potentially interact (intentionally or else), and their relevant properties, during the time spanned by the environment and nothing more. For example, if the environment is a spatio-temporal region with colored objects in it and the agent belongs to a type that detect physical objects but not their color, then the objects in the environment of this agent do not have color properties (the environment is neutral on their color). Similarly, if the agent cannot move (e.g. a robotic arm in an industrial site), its environment is limited to the space the arm can reach either directly or via other devices (e.g., a detachable component) when present. If at a certain point a graspable tool is accessible to the robot, the tool becomes part of the agent's environment which, in turn, expands to the space reachable by the arm grasping the tool (assuming the robot has a suitable grasp capability)[1]. If the agent can move and the environment has duration t,

[1] Agent's environment, capability and affordance are clearly intertwined notions.

the agent's environment includes every object that the agent can potentially interact with in that interval of time. Note that I am using a general notion of *interaction*: it refers to any action the agent performs or undergoes due to the environment, including balancing gravity and collecting information (like when it detects the presence of another object). It follows that two agents of different types most likely have two different agent's environments even though they are in the same general environment, i.e., spatio-temporal region.

It might not be trivial to map existing notions of behavior into this ontological characterization. For instance, in robotics one talks of reactive behavior, observed behavior and obstacle avoidance behavior. Reactive behavior is a subtype of agent's behavior where the behavior depends on past and actual interactions (modulo reaction time). Observed behavior restricts behavior to the features that an observer of that type can possibly detect (an observer-dependent reduction). Obstacle avoidance behavior is a combination of reactive behavior and function execution, the latter needs an ontological functional module to be analyzed.

Here I cannot discuss the very notion of agent or go further into details. It should be clear that the framework is conceived as a module for ontologies, and that it classifies every actual (and possible or hypothetical) behavior by considering it a relational property of agents in an environment. Among the behavior specializations to include, are the internal vs. external behavior (focus on the agent's internal components vs. the external objects in the environment) which is a restriction on the environment, and physical vs. information behavior (behavior in the physical vs. the information space), which is a restriction on interaction types. Fortunately, there is already some work on these distinctions in applied ontology.

Conclusions

The paper argued that collaboration among biological and artificial agents requires a broader, yet sound and reliable, notion of behavior. This can be achieved by integrating and unifying research in behavioral studies with that in engineering, robotics and human-robotics interaction. A starting point has been set showing how to take advantage of results in the domain of applied ontology.

References

1. Borgo, S. (2019) Ontological challenges to cohabiting with self-taught robots. *SWJ*.
2. Guarino, N., Oberle, D., Staab, S. (2009). What is an ontology? In *Handbook on ontologies*, pages 1–17. Springer.
3. Lewis, D. (2008). *Convention: A philosophical study.* John Wiley & Sons.
4. Menzel, P., d'Aluisio, F. (2000) *Robo sapiens: Evolution of a new species* MIT Press.
5. Nolfi, S. et al. (2016) Evolutionary robotics. *Handbook of Robotics*, 2035–68 Springer.
6. Pinto, L., Gupta, A. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In *2016 IEEE ICRA* (2016), pp. 3406–3413.
7. Lazaridou, A., Peysakhovich, A., Baroni, M. Multi-agent cooperation and the emergence of (natural) language, 2016. preprint arXiv:1612.07182.