Towards Institutions for Mixed Human-Robot Societies

Robotics Track

Stevan Tomic
Örebro University
Örebro, Sweden

Alicja Wasik
Institute for Systems and Robotics
Lisbon, Portugal

Pedro U. Lima
Institute for Systems and Robotics
Lisbon, Portugal

Alcherio Martinoli
Ecole Polytech Fédérale Lausanne
Lausanne, Switzerland

Federico Pecora
Örebro University
Örebro, Sweden

Alessandro Saffiotti
Örebro University
Örebro, Sweden

ABSTRACT

We report an exploration into normative reasoning for robots in human societies using the concept of institutions.

KEYWORDS

Normative agents; Social robots; Cognitive robotic; Human Robot Interaction; Multi-Robot Systems

ACM Reference Format:
Stevan Tomic, Alicja Wasik, Pedro U. Lima, Alcherio Martinoli, Federico Pecora, and Alessandro Saffiotti. 2018. Towards Institutions for Mixed Human-Robot Societies. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

There is little doubt that future robots will need to comply to the norms and conventions that humans use. One could hard-wire such compliance by programming ad-hoc behaviors into our robots. In this paper we take a different, model-based approach: we codify generic norms into reusable structures, called Institutions, and then ground these institutions into specific domains. This is in accordance to North's view that institutions constitute "the human devised constraints that shape social interaction" [3]. We show how our approach can be used to introduce normative aspects into robot planning and control, thus allowing robots to participate in mixed human-robot societies that adhere to human-defined norms. We also show how the same institution can be used in different situations by only changing the grounding relation.

2 FORMAL MODEL

The full framework for grounded reasoning with institutions is presented in [4]. In this section we only introduce the parts of that framework that are relevant to this paper.

Our institutions encapsulate a collection of *norms* together with the roles: $Roles = \{role_1, \ldots, role_m\}$, actions: $Acts = \{act_1, \ldots, act_k\}$ and artifacts: $Arts = \{art_1, \ldots, art_a\}$ that they refer to. Examples of *Roles* are a 'tour-guide' in a museum, or a 'customer' in a store. *Arts* are the relevant objects, e.g., a 'painting' or a 'cash-register'. *Acts* are actions that can be performed, e.g., 'describe' a painting or 'buy' a milk box. Thus, for example, a "guided exhibition" can be modeled by an institution that includes guides, exhibits, and a norm stating that guides must describe exhibits. We define norms to be predications over statements, where a statement is a relation between a subject, a predicate and an object:

Definition 1. A *norm* has the form $q(trp^*)$, where q is a qualifier and trp is a triple of the form:

$$trp \in Roles \times Acts \times (Arts \cup Roles)$$

Qualifiers can be unary relations like must or must – not, or n-ary ones like inside or before. For example, must can be used to represent the obligation "A tour-guide *must* describe a painting": must (⟨TourGuide, Describe, Painting⟩) . While the binary qualifier in_front_of can specify a spatial constraint on the location of an action, as in "Describing should be performed *in front of* a painting": in_front_of (⟨TourGuide, Describe, Painting⟩). Binary qualifiers can express relations between statements, e.g., temporal relations such as before or during. Institutions put all the above elements together:

Definition 2. An *institution* is a tuple

$$\mathcal{I} = \langle Arts, Roles, Acts, Norms \rangle.$$

An institution is an abstraction, which can be instantiated in concrete systems that are physically different. For instance, the same "guided exhibition" institution can be used to regulate behaviors of agents in different museums, irrespective of these agents being humans, robots, or a combination of both. We call such a concrete system a *domain.*

Definition 3. A *domain* is a tuple $\mathcal{D} = \langle A, O, B, F, R \rangle$, where

- A is a set of *agents,*
- O is a set of *physical entities,*
- B is a set of *behaviors,*
- $F \subseteq A \times B \times (O \cup A)$ is a set of *affordances.*
- R is a finite set of *state variables.*

The set A can include humans (e.g., John), robots (e.g., nao), or both. B is the collection of all behaviors that these humans or robots can perform, like moveTo or speak. O are objects in the domain, like EntranceDoor, or StarryNight. The F relation indicates which agents can execute which behaviors and on which objects, e.g.,

$$\{ \langle \text{John}, \text{moveTo}, \text{Irises} \rangle, \langle \text{nao}_1, \text{speak}, \text{StarryNight} \rangle \}$$

Definition 4. Given an institution $\mathcal{I}$ and a domain $\mathcal{D}$, a *grounding* of $\mathcal{I}$ into $\mathcal{D}$ is a tuple $\mathcal{G} = \langle \mathcal{G}_A, \mathcal{G}_B, \mathcal{G}_O \rangle$, where: $\mathcal{G}_A \subseteq Roles \times A$ is a

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

role grounding, $\mathcal{G}_B \subseteq Acts \times B$ is an *action grounding*, $\mathcal{G}_O \subseteq Arts \times O$ is an *artifact grounding*.

Grounding plays an important role in our framework, by establishing the relation between roles and agents, generic actions and behaviors of agents, and institution artifacts and physical objects. It provides the key to reuse the same abstract institution to describe or regulate different systems. For example, the "guided exhibition" institution can be grounded in different physical museums (domains) using different $\mathcal{G}$'s, but another $\mathcal{G}'$ may be used to ground it in a university where a professor shows a lab to visitors.

Grounding allows us to give norms a concrete semantics in terms of execution in a physical domain. For example, the semantics of the norm $\text{must}\,((\text{tour} - \text{guide}, \text{describe}, \text{painting}))$ is given by all executions where any agent a that *grounds* the role *tour-guide* executes at least once a behavior b that *grounds* the action *describe*. In this way we can tell if a given physical execution complies to all norms in an institution.

3 CASE STUDY

We show how our institutional framework can be used to govern a physical system with robots and humans.

The scenarios consists of two phases. In the first phase, governed by Guidance institution, a group of human visitors is guided by a number of agents to a large room. The formation includes one leader, the visiting humans, and the followers. The task of the latter is to contain the visitors within close proximity behind the leader. Phase two, govern by Tutoring institution, takes place in the large room, where two cooperative agents take the visitors on a tour of the room, explaining the objects in the room to the visitors. One of the agents describes each object, while the other moves around the classroom, showing the objects to the visitors. The Guidance institution is defined as follows: $Guidance = \langle Arts_G, Roles_G, Acts_G, Norms_G \rangle$, where:

$$
\begin{aligned}
Roles_G &= \{\text{Leader, Follower, Visitor}\} \\
Acts_G &= \{\text{Guide, AssistGuidance}\} \\
Arts_G &= \{\text{Destination}\} \\
Norms_G &= \{\text{must}\,((\text{Leader, Guide, Destination}))\,, \\
&\quad \text{must}\,((\text{Follower, AssistGuidance, Leader})) \\
&\quad \text{use}\,((\text{Leader, Guide, Destination})) \\
&\quad \text{use}\,((\text{Follower, AssistGuidance, Leader})) \\
&\quad \text{keep\_distance}\,((\text{Follower, AssistGuidance, Leader}))\,, \\
&\quad \text{keep\_distance}\,((\text{Follower, AssistGuidance, Visitor}))\,, \\
&\quad \text{keep\_distance}\,((\text{Visitor, Follow, Leader})) \\
&\quad \text{while}\,((\text{Leader, Guide, Destination})\,, \\
&\qquad\quad (\text{Follower, AssistGuidance, Leader}))\}
\end{aligned}
$$

The domain is given as set of agents: $A = \{\text{mbot}_{03}, \text{mbot}_{11}, \text{nao}, \text{child}_1, \ldots\}$, behaviors: $B = \{\text{MoveOnTrajectory}, \text{MoveInFormation}, \text{PerformDialog}, \text{ShowObject}\}$, and objects: $O = \{\text{largeRoom}, \text{entrance}_1, \text{yellowSphere}, \ldots\}$. The affordances are:

$$
\begin{aligned}
F = \{ &\left(\text{mbot}_x, \text{MoveInFormation}, \text{mbot}_y\right), \\
&(\text{mbot}_x, \text{MoveOnTrajectory}, \text{entrance}_1)\,, \\
&(\text{mbot}_x, \text{ShowObject}, \text{object}_z), (\text{nao}, \text{PerformDialog}, \text{object}_z)\}
\end{aligned}
$$

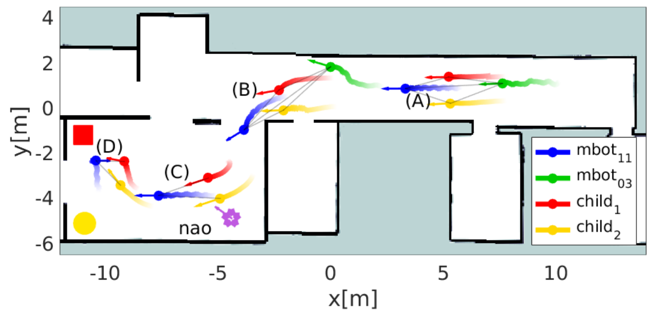

Figure 1: Estimated trajectories of the humans and the robots during an execution of the first scenario.

where we use the subscripts $(\cdot)_x$, $(\cdot)_y$ and $(\cdot)_z$ to abbreviate several affordances that involve agent/object $(\cdot)$. Human behaviors, used by the Visitors, are represented using a realistic model of a pedestrian motion [1]. Robots behaviors are described as follows:

(1) MoveOnTrajectory: plan and follow a path in a known environment [5].
(2) MoveInFormation: assist the Leader in guiding the Visitors [6].
(3) PerformDialog: perform a speech act.
(4) ShowObject: Navigate to a grounded waypoints.

Experiments were performed in a realistic simulator [2]. Behavior control was distributed, and the planner was centralized. We run experiments on two scenarios using two different groundings. In the first scenario, the following grounding was used:

(1) *The Guidance Institution* with $\mathcal{G}_{\text{Guidance}}$:

$$
\begin{aligned}
\mathcal{G}_A = &\ \{(\text{Leader, mbot}_{11}), (\text{Follower, mbot}_{03}), \\
&\ (\text{Visitor, child}_1), (\text{Visitor, child}_2)\} \\
\mathcal{G}_B = &\ \{(\text{Guide, MoveOnTrajectory}), \\
&\ (\text{AssistGuidance, MoveInFormation})\} \\
\mathcal{G}_O = &\ \{(\text{Destination, entrance1})\}
\end{aligned}
$$

In the second scenario, the same norms were followed by different robots on different artifacts to illustrates re-usability of institutions and the role of grounding. Figure 1 shows an execution of the first scenario. A video showing executions of both scenarios is available at: https://tinyurl.com/yanurn3j. An additional video shows execution on real robots: https://tinyurl.com/ycwxerfv.

Our next step will be to validate our approach in real settings and to perform a user-based evaluation aimed at verifying the hypothesis that following human-defined norms improves user's acceptance of robots.

REFERENCES

[1] D. Helbing and P. Molnár. 1995. Social force model for pedestrian dynamics. *Physical Review E* 51, 5 (1995), 4282–4286. http://arxiv.org/abs/cond-mat/9805244
[2] O. Michel. 2004. Cyberbotics Ltd. Webots: Professional Mobile Robot Simulation. *International Journal of Advanced Robotic Systems* 1, 1 (2004), 5.
[3] Douglass C North. 1990. *Institutions, institutional change and economic performance.* Cambridge university press.
[4] Tomic Stevan. 2018. *Institutions, Norms, and Robots.* Ph.D. Dissertation. Orebro University. Forthcoming.
[5] R. Ventura and A. Ahmad. 2015. Towards Optimal Robot Navigation in Urban Homes. In *RoboCup 2014: Robot World Cup XVIII.* Springer, 318–331.
[6] A. Wasik, J.N. Pereira, R. Ventura, P.U. Lima, and A. Martinoli. 2016. Graph-Based Distributed Control for Adaptive Multi-Robot Patrolling using Local Formation Transformation. In *IEEE Int Conf on Intelligent Robots and Systems.* 1721–1728.