Learning Normative Behavior through Abstraction: Semantics Details and Hyper-Parameters used for Training

Version 0.1*

The purpose of this document is to enable the reproducibility of reported experiments in the paper "Learning Normative Behavior through Abstraction" by S.Tomic, F. Pecora and A. Saffiotti. The document provides a more detailed overview of norms semantics and lists all training parameters used for the training.

Norms Semantics Used in the Experiments

Semantics depends on grounding. Thus let's define following sets:

$$A_{role} = \mathcal{G}_A(role), B_{act} = \mathcal{G}_B(act), O_{art} = \mathcal{G}_O(art)$$

That is, A_{role} is the set of all agents $a \in A$ ground by $role$, B_{act} is the set where all behaviors $b \in B$ are grounded by act, and O_{art} is the set of all objects $o \in O$ grounded by art.

The norms semantics is a subset of all trajectories that fulfill provided formula listed below.

MustUse

$$[\![\text{mustUse}((role, act, art))]\!]_F \equiv$$
$$\{(I, \tau) \mid \forall a \in A_{role}.\exists (b, o, t) \in B_{act} \times O_{art} \times I :$$
$$\text{active}(b, a)(\tau(t)) = \top \wedge (\text{active}(b, a)(\tau(t)) = \top \implies \text{usedObj}(b, a)(\tau(t)) = o))\}$$

*This version contains only semantics as logical formulas. The next version should include source code for each semantics.

Activation of the grounded behavior in the grounded agent at arbitrary time t has to be true ($\text{active}(b, a) = \top$), and if it is true then used object has to be one grounded by the art in the norm ($\text{usedObj}(b, a) = o)$)) in time t.

MustAt

$$
\begin{aligned}
&[\![\, \text{mustAt}\,((role, act, art))]\!]_F \equiv \\
&\{(I, \tau) \mid \forall a \in A_{role}.\exists (b, o, t) \in B_{act} \times O_{art} \times I : \\
&\quad \text{active}(b, a)(\tau(t)) = \top \wedge (\text{active}(b, a)(\tau(t)) = \top \implies \text{near}(o, a)(\tau(t)) = \top)\}
\end{aligned}
$$

Activation of the grounded behavior in the grounded agent has to be true ($\text{active}(b, a) = \top$) at time t and if it is true then the position of behavior execution for the grounded agent has to be the same as position of the grounded object ($\text{active}(b, a) = \top \implies \text{near}(o, a) = \top$) at time t.

Before

The purpose of 'before' semantics is to indicate if something happened in the desired temporal order, that is, does a certain value in state-variable ρ_1 has happened before the value in state-variable ρ_2. As such, semantics of 'before' relates values of two such variables in the temporal dimension. Thus, if one is interested weather an agent was in a certain place before some other agent, 'before' semantics can relate position state-variables of two agents by comparing their values. However, if one is interested weather an agent has used the object before another agent then the state-variable 'usedObj' have to be in temporal relation. This will create two separate semantics of the concept of 'before', which is a drawback. The temporal relations can be further abstracted (not depending on such use-cases), but it is not the topics of this document.

Here two variants of before semantics are used, one defining temporal relation between 'useObj' variable and both 'active' and 'near' variables, and the other concerning only 'active' and 'near' state-variables.

BeforeUseAt

$$
\begin{aligned}
&[\![\,\text{before}\,((role_1, act_1, art_1), (role_2, act_2, art_2))]\!]_F \equiv \\
&\{(I,\tau) \mid \forall (a_1,a_2) \in A_{role_1} \times A_{role_2}, (b_1,b_2) \in B_{act_1} \times B_{act_2}, (o_1,o_2) \in O_{art_1} \times O_{art_2}, \\
&\ \forall (t_1,t_2) \in I \times I, \\
&\ (\text{usedObj}(b_1,a_1)(\tau(t_1)) = o_1 \wedge \text{active}(b_2,a_2)(\tau(t_2)) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2)) = \top \\
&\quad \Longrightarrow\ t_1 < t_2) \wedge \\
&\ (\forall (a_1,a_2) \in A_{role_1} \times A_{role_2}, (b_1,b_2) \in B_{act_1} \times B_{act_2}, (o_1,o_2) \in O_{art_1} \times O_{art_2}, \\
&\ \exists (t_1,t_2) \in I \times I : \\
&\quad \text{usedObj}(b_1,a_1)(\tau(t_1)) = o_1 \wedge \text{active}(b_2,a_2)(\tau(t_2)) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2) = \top))\}.
\end{aligned}
$$

This semantics concerns trajectories in which a grounded agent using an object ($\text{usedObj}(b_1,a_1) = o_1$) has to precede activation of another grounded agent $\text{active}(b_2,a_2) = \top$, near the grounded object ($\text{near}(o_2,a_2) = \top$) and such usages and activations has to exists ($\exists (t_1,t_2) \in I \times I : \dots$).

The *violation* semantics is similar:

$$
\begin{aligned}
&[\![\,\text{before}\,((role_1, act_1, art_1), (role_2, act_2, art_2))]\!]_V \equiv \\
&\{(I,\tau) \mid \forall (a_1,a_2) \in A_{role_1} \times A_{role_2}, (b_1,b_2) \in B_{act_1} \times B_{act_2}, (o_1,o_2) \in O_{art_1} \times O_{art_2}, \\
&\ \forall (t_1,t_2) \in I \times I, \\
&\ (\text{usedObj}(b_1,a_1)(\tau(t_1)) = o_1 \wedge \text{active}(b_2,a_2)(\tau(t_2)) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2)) = \top \\
&\quad \Longrightarrow\ t_1 >= t_2) \wedge \exists t_2' \in I : \\
&\ \text{active}(b_2,a_2)(\tau(t_2')) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2')) = \top \wedge t_2 = t_2'\}
\end{aligned}
$$

Here semantics checks the same conditions of state-variables but the relation sign is reversed ($t_1 >= t_2$), that is, if the order is different the violation semantics is satisfied. If only the second condition happens $\text{active}(b_2,a_2) = \top \wedge \text{near}(o_2,a_2) = \top$ then the violation condition is also satisfied. This can be used as a signal to reinforcement learning agents as soon as the temporal norm is violated.

BeforeAtAt

The semantics of 'before' above constraints trajectories regarding the temporal dimension of usage and position of behaviors and objects. Semantics

'BeforeAtAt' constraints trajectory regarding the temporal relation of spatial state-variables concerning activation and the position of agents:

$$
\begin{aligned}
&[\![\,\text{before}\,((role_1, act_1, art_1), (role_2, act_2, art_2))]\!]_F \equiv \\
&\{(I,\tau) \mid \forall (a_1,a_2) \in A_{role_1} \times A_{role_2}, (b_1,b_2) \times B_{act_1} \times B_{act_2}, (o_1,o_2) \times O_{art_1} \times O_{art_2}, \\
&\ \forall (t_1,t_2) \in I \times I, \\
&\ (\text{active}(b_1,a_1)(\tau(t_1)) = \top \wedge \text{near}(o_1,a_1)(\tau(t_1)) = \top \wedge \\
&\ \text{active}(b_2,a_2)(\tau(t_2)) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2)) = \top \implies t_1 < t_2) \wedge \\
&\ (\forall (a_1,a_2) \in A_{role_1} \times A_{role_2}, (b_1,b_2) \times B_{act_1} \times B_{act_2}, \exists (t_1,t_2) \in I \times I : \\
&\ \text{active}(b_1,a_1)(\tau(t_1)) = \top \wedge \text{near}(o_1,a_1)(\tau(t_1)) = \top \wedge \\
&\ \text{active}(b_2,a_2)(\tau(t_2)) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2)) = \top)\}.
\end{aligned}
$$

The activation of behavior ($\text{active}(b_1, a_1) = \top$) in a position near the specific object ($\text{near}(o_1, a_1) = \top$), both at t_1, should precede activation of another grounded behavior ($\text{active}(b_2, a_2) = \top$) near another specified object ($\text{near}(o_2, a_2) = \top$) at t_2, and such activations should exists ($\exists (t_1, t_2) \in I \times I : \ldots$).

Equals

$$
\begin{aligned}
&[\![\,\text{equals}\,((role_1, act_1, art_1), (role_2, act_2, art_2))]\!]_F \equiv \\
&\{(I,\tau) \mid \forall (a_1,a_2) \in A_{role_1} \times A_{role_2}, (b_1,b_2) \times B_{act_1} \times B_{act_2}, (o_1,o_2) \times O_{art_1} \times O_{art_2}, \\
&\ \forall (t_1,t_2) \in I \times I, \\
&\ (\text{active}(b_1,a_1)(\tau(t_1)) = \top \wedge \text{near}(o_1,a_1)(\tau(t_1)) = \top \wedge \\
&\ \text{active}(b_2,a_2)(\tau(t_2)) = \top \wedge \text{near}(o_2,a_2)(\tau(t_2)) = \top \implies t_1 == t_2)\}.
\end{aligned}
$$

The activation of behavior ($\text{active}(b_1, a_1) = \top$) in a position near the specific object ($\text{near}(o_1, a_1) = \top$), both at t_1, should be same in time as the activation of another grounded behavior ($\text{active}(b_2, a_2) = \top$) near another specified object ($\text{near}(o_2, a_2) = \top$) at t_2, so that $t_1 == t_2$.

Hyper-parameters Used for Learning in All Experiments

The following hyper-parameters are used:

```
default:
trainer: ppo
batch_size: 1024
beta: 5.0e-3
buffer_size: 10240
epsilon: 0.2
gamma: 0.99
hidden_units: 128
lambd: 0.95
learning_rate: 3.0e-4
max_steps: 5.0e4
memory_size: 256
normalize: false
num_epoch: 3
num_layers: 2
time_horizon: 64
sequence_length: 64
summary_freq: 1000
use_recurrent: false
use_curiosity: false
curiosity_strength: 0.01
curiosity_enc_size: 128

#ALL BRAINS (normal and abstracted):
AgentBrain:
beta: 6e-3
max_steps: 4.0e6
hidden_units: 256
num_layers: 3
time_horizon: 1024
```