Inference of the Intentions of Unknown Agents in a Theory of Mind Setting

Michele Persiani*1[0000−0001−5993−3292] and Thomas Hellström*1[0000−0001−7242−2200]

Department of Computing Science, Umeå University, Umeå, Sweden
{michelep,thomash}@cs.umu.se

Abstract. Autonomous agents may be required to form an understanding of other agents for which they don't possess a model. In such cases, they must rely on their previously gathered knowledge of agents, and ground the observed behaviors in the models this knowledge describes by theory of mind reasoning. To give flesh to this process, in this paper we propose an algorithm to ground observations on a combination of priorly possessed Belief-Desire-Intention models, while using rationality to infer unobservable variables. This allows to jointly infer beliefs, goals and intentions of an unknown observed agent by using only available models.

1 Introduction

An important aspect emphasized in recent research on intent recognition is that the actor agent, whose intention should be found, and the observer agent, who attempts to infer the intention, may be using different models to represent each other. In this decoupled setting, the observer must form a model of the actor's decision-making process in order to understand its actions. This creation of another agent's model is commonly referred to as theory of mind reasoning [3], or a first-order theory of mind. However, commonly this model is given a priori to the observer [5, 3, 11], or is assumed to be equivalent to the one it is already using [8, 12, 17]. Such assumptions may work well in hand-crafted or simple domains but is unrealistic if the agents are heterogeneous and autonomous. In such cases, they should rather build models of each other through observations and interaction.

The algorithmic creation of a theory of mind that goes beyond simple controlled experiments is still a hard problem [4]. While most research focuses on reconstructing internal beliefs from observations, an additional difficulty seldom addressed is that the observer cannot know the symbols and schemas that the actor is using to create its beliefs or its deliberation model. The only symbols available to the observer are those it itself possesses, and the models it can form through observations must be a function of those symbols alone. In this paper we address the following questions: how can the observer realize the actor's model based on schemas that are available to him? And, how can the observer infer the actor's intentions based on those?

To answer these questions, and in agreement with earlier work, we propose that the only key assumption we need is that the actor is intentional, and acts rationally to pursue its goal (commonly, this is referred to as being subject to the *principle of rational action* [1]). Therefore, the models that fulfil these requirements are the candidate hypotheses for the true model of the actor. This resulting space of hypotheses describes what must be true or false if the agent is performing intentionally, both in terms of beliefs and how these beliefs and their representations are combined to form an action schema. Everyone of these models could describe different hypothesis, yet in all of them the actor is explained as being intentional. They are all possible hypothesis for the true world in which the actor is being intentional and therefore we here propose that they are *equivalent* to the true actor's model when trying to understand what are its beliefs and goals. Since these hypothesized models could be symbolically heterogeneous, they may form multiple descriptions of the actor's intention because of their different symbolic forms. Therefore, in order to find the models in which the actor is being intentional, we can project an assumed optimality of observations over the space of possible models, after which the valid models underlying rationality are those that allow to explain the observations as being optimally directed towards a goal. We define this class of models as the equivalent class of rational models.

We propose that the observer can generate an initial guess of this class of models (possibly starting fully unspecified) using its known schemas, then refining it by maximizing the rationality expressed by the observations. This inference is possible from the observer perspective, and doesn't require the true symbols the actor is using to represent its world. Having the set of rational models, we find the probability of a certain goal, or predicate in the agent's belief.

In this paper we propose a novel algorithm for constructing models of an observed agent, based on the maximization of rationality in the observations. The proposed method utilizes the Planning Domain Description Language (PDDL), that allows us to easily perform tests on arbitrary domains. We extend earlier theory of mind and intent recognition formalizations by simultaneously considering multiple candidate models. In Section 2 we describe how the proposed methods fit in the current literature. In Section 3 we describe our proposed method to find an agent's equivalence class of rational models from observations, followed by a description of how we implemented it using PDDL in Section 4. In Section 5 we provide a simple illustrative example and experimental results on a joint belief and goal inference task, performed on several standard domains. Finally, Sections 6-7 describe the current limitations, proposed future work, and conclusions.

2 Background and Related Work

Intent recognition is the algorithmic task of finding an agent's intention using some observations as evidence. In planning contexts where agents can move in an artificial world and take decisions, intentions can be understood as an agent's plan of actions and/or desired goal state. As also previously shown, in this setting intent recognition can be realized by goal or plan recognition techniques [15].

Recent research suggests to complement intent recognition with theory of mind reasoning. In its context, intentions form relevant parts of an agent's *state of mind* [10]. An example which motivates the utilization of theory of mind for intent recognition is the following. Let's suppose that an AI autonomously managing a building, in which it is embodied, attempts to infer the goal destination of a person walking an hallway. Clearly, the knowledge in terms of the state of the building is largely different between AI and person. Since the AI can gather a lot of data from its sensors, its instantaneous state is very rich in details e.g. knows who is in the building and where, which doors are open, etc. In this setting, computing the person's intention using the AI's belief is wrongly assuming that the person possesses the same amount of information. Therefore, to correctly make predictions, the AI should first estimate what are the person's beliefs, to then perform intent recognition based on those. i.e. it must form a theory of mind of the person that is focused on his belief about the building. Crucially, this allows to perform tests of *false belief.* For example, supposing that the AI knows that a door is closed, observing a person going towards the door without before taking its key allows to infer that the person has a false belief of the door being open.

An important point often only scratched in the literature is about the prior models that are provided to the agents doing inference. Often, these models are assumed to be completely known such as in [3], where the authors use Bayesian inference on POMDPs to jointly infer an agent's beliefs and goal while navigating a grid environment with multiple possible goals. The authors show that intent recognition using theory of mind reasoning forms predictions that are comparable to humans predictions. However, in their work a model of how the actor perceives, can move, etc. is explicitly required. Rather, in this paper we utilize the class of rational models that are induced by the rationality in the observations, which is the only assumed property of the actor agent. We consider multiple candidate models for the actor rather than a single one. Additionally, these models come from priors internally possessed by the observer, and another relevant divergent point is that we don't require a true model of the actor agent (and in particular its observation function). While in past research on theory of mind reasoning computing an observation function of the actor was considered positively grounded in folk psychology to the mechanism of *spontaneous perspective taking*, recent research is criticizing the position [6] by arguing that there is not enough evidence to claim that humans consistently do perspective-taking in interactions, as well as to describe how humans infer others visual perspectives. Computing an observation function of a robot's human collaborators has been shown to be feasible in highly controlled environments [7], however, in this paper we assume for the observation function of the actor agent to be unknown, with the observations being gathered from the observer's observation model only.

In [16] the authors propose a neural network which learns multiple species of agents moving in a gridworld, each of which described by a POMDP. Their *Machine Theory of Mind* shows that it is possible to memorize multiple models of agents, to then infer posterior distributions about their beliefs and intentions from observations. Despite the diversity of the produced agents, the number of classes of agents is large but still limited by the dataset. An important drawback of this approach is that the dataset must enumerate all of the possible models and observation functions of the POMDP agents, which quickly becomes intractable in complex scenarios.

Other relevant background research is in *Epistemic Plan Recognition* [19], that is a formalization for planning and plan recognition problems in multi-agent settings that explicitly takes into account observers and their beliefs. And the seminal work in plan and goal recognition of [18], over which part of our discussion on rationality is based on.

3 Method

We model the actor agent as a Belief-Desire-Intention (BDI) agent, and the observer agent as an intent recognition agent that infers the best candidates of the actor's BDI components using gathered observations. The BDI architecture [9] is a common framework to model agents. In a BDI agent, the beliefs comprise what is true for the agent. Desires correspond to possible goal states, while intentions are plans of actions, consistent with the beliefs, and obtained through a deliberation model, in which the agent commits to fulfil its desires. In this setting we model intentions as a function of the beliefs and a goal. Intentions are consistent with the beliefs, and could, for example, be generated by deliberation cycles computed at every time step of the agent [2]. Therefore, we assume that whenever the agent model is known, knowing also the agent's goal and beliefs is sufficient to compute its intention, for example by using a plan library or a forward planning procedure. Intent recognition can therefore be seen as a two-step process. First, the observer evaluates belief-goal pairs to find the one best matching the observations. Second, the intentions are found by simulating this selected actor model.

In the space of possible models that the observer uses to explain the actor's actions as intentional we refer to the subset of models that preserve rationality as the actor's equivalence class of rational models. They represent the class of models that are equivalent in preserving the actor's optimality toward a possible goal, by capturing what must be true if the agent is behaving intentionally.

Taking the previously given example of the person walking the hallway, the observations could be explained in a number of ways, each of which ground the person's actions in a different BDI model, such as gridworld model (the person seeks to reach a tile in the world), social model (the person seeks to reach a person), a combination of those or others. Raw observations are grounded in the selected set of models, in which the corresponding intentions are evaluated.

Rather than using a single BDI model, the observer may combine multiple models to understand the actor's actions. This is equivalent to refining an initial distribution of models $P(\Xi;\theta)$ towards the agent's class of rational models $P(\Xi;\theta_R)$, where $\xi = (b, g, a) \sim \Xi$ is a sampled instance which includes a candidate agent model, goal and belief. For simplicity, we assume that all models in Ξ are compatible with the underlying sources of data, or alternatively, the observer considers only the models that are possible given its context.

The inference process of the observer is represented in a graphical form in Figure 1. The observer has a model of the actor as a joint probability distribution of beliefs, goals and deliberation models $P(\Xi_{act}^{obs}) = P(B, G, A)$. Every instance $\xi \sim P(\Xi)$ contains a fully specified, candidate description of the actor's BDI state. For example, a belief $b \in B$ could be described with a set of truth predicates, a goal $g \in G$ as the desired belief state. A deliberation model $a \in A$ is an action schema as we will later show. We consider an intention $\pi \in \Pi_{act}^{obs}$ as a committed plan recipe consistent with a deliberation model, together with the goal it attempts to achieve. For a particular $\xi = (b, g, a)$ the instantiated intentions are those plans that are consistent with (b, a) and that fulfills g.

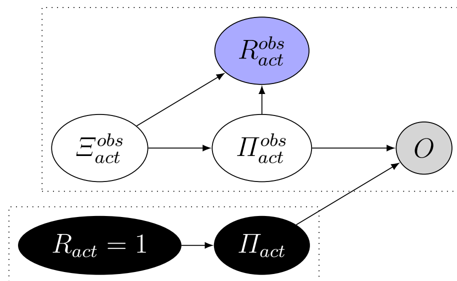

Fig. 1: Graphical model describing the variables involved in intent recognition and their connection in a theory of mind context. $\mathbf{\Xi_{obs}^{act}}$: Inferred agent model, $\mathbf{\Pi_{obs}^{act}}$: Inferred intention, **O**: Observations, **R**: Rationality. The figure highlights that in order to infer intentions the observer must beforehand internally model the actor agent. The only assumption we make about the actor is that its intentions are rational.

As they unfold, intentions produce observations that can be gathered by the observer. For an agent instance ξ, its candidate intentions are inferred using the set of observations $o = \{o_1, ..., o_n\} \in O$ that, once grounded, describe the effects that the actor's actions had on the world as described by ξ. Therefore, inferred intentions must be consistent both with the considered actor's models and the gathered observations. This is highlighted in the following probability distribution from the Baysian netwrok in Fig. 1:

$$P(\Pi|O)P(O) = P(O|\Pi) \sum_{\Xi} P(\Pi|\Xi)P(\Xi) \propto \sum_{\Xi} P(\Pi|\Xi, O)P(\Xi) \qquad (1)$$

where Π is the random variable of possible plans, Ξ of possible BDI instances, and O of possible observations. Eq. 1 must provide high likelihoods for intentions that are consistent with both the observations and considered models. This can be achieved by setting $P(O = o|\Pi = \pi) > 0$ only if $o \in \pi$, and $P(\Pi = \pi|\Xi = \xi) > 0$ only if π is a plan consistent with ξ. Additionally, since the actor is assumed to be intentional, $P(\Pi = \pi | \Xi = \xi)$ should reflect the rationality of π in ξ. In Section 4 we show how it can be implemented using the Planning Domain Description Language.

3.1 Maximization of rationality

The key idea driving intentions is that they must be rational. Therefore, in order to find the actor's intentional model the observations should be interpreted in a way that explains them as rational, or, in our case, optimal in some possible ξ. Following this argument we define rationality as a measurable property of observation-instance pairs, $R(o, \xi)$, defined as the expected rationality that the instance's intentions have while being constrained to be consistent with the observations:

$$R(o, \xi) = E_{\pi \sim P(\Pi | \Xi = \xi, O = o)}[R(\xi, \pi)] \tag{2}$$

where $R(\xi, \pi)$ is the rationality of a specific plan computed in ξ. In agreement with the principle of rational action, we set $R(\xi, \pi)$ to be in function of the optimality of π in achieving ξ's goal. This optimality measure for a plan π can be expressed as:

$$R(\xi, \pi) = \exp\{|\pi_{opt,\xi}| - |\pi|\} \tag{3}$$

where $\pi_{opt,\xi}$ is an optimal plan for ξ. Since $|\pi_{opt,\xi}| \leq |\pi|$, $R(\xi, \pi)$ has a value of 1 if the observations are along an optimal plan, a value between 0 and 1 whenever the observations belong to a sub-optimal plan. The observation o is rational in ξ if the intentions produced by $P(\Pi | \Xi = \xi, O = o)$ fulfill the principle of rational action, that is, they likely correspond to the optimal plans obtainable in ξ. When matched against all possible instances from the distribution of models $P(\Xi)$, an expected rationality of the observations is obtained as:

$$R(o) = E_{\xi \sim P(\Xi)}[R(\xi, o)] = E_{\pi \sim P(\Pi | \Xi = \xi, O = o), \xi \sim P(\Xi)}[R(\xi, \pi)] \tag{4}$$

Our proposed method for finding plausible agent models is to maximize the expected rationality of the observations $R(o)$. This is because, as we introduced in Section 1, we aim to search the model space to find instances expressing a rational behavior. Therefore, we are interested in finding the planning instances that maximize the degree of rationality $R(\xi, \pi)$ of intentions consistent with the observations, i.e. that also have a high likelihood $p(\pi | \xi, o)$. At the end of the optimization process, sampling from the resulting distribution yields planning instances in which the observations are contained in maximally rational intentions. Therefore, after training, $P(\Xi; \theta_R)$ captures a distribution of BDI models that explain the observed agent behavior as rational. It is the agent's equivalence class of rational models. In order to train $P(\Xi; \theta)$ we start by considering the expected value of rationality of a sequence of observations o and the parameters θ_R that maximize $R(o)$:

$$R(o) = E_{\xi \sim P(\Xi;\theta)}[R(\xi, o)] = \sum_{\Xi} R(\xi, o)p(\xi;\theta),\ \theta_R = \underset{\theta}{\operatorname{argmax}} \sum_{\Xi} R(\xi, o)p(\xi;\theta) \tag{5}$$

This maximization is difficult for two main reasons. The space of planning instances defined by $P(\Xi;\theta)$ can be very large, and in the general case $R(o)$ is non differentiable since it requires to compute plans through e.g. a planner. To overcome this issues we propose the following Expectation-Maximization (E-M) procedure based on sampling, which avoids to compute the derivative of the rationality function.

3.2 E-M importance sampling

To speed up the E-M algorithm we introduce an importance sampling buffer $P_R(\Xi)$ with limited capacity that holds past generated planning instance with high rationality. By using the memory buffer, planning instances are sampled using probabilities based on their rationality rather than on the current parameters value of $P(\Xi;\theta)$. Instances sampled during the E-step are sampled from this buffer rather than being freshly generated using $P(\Xi;\theta)$. When sampling instances from the buffer we have:

$$\xi \sim P_R(\Xi), p_R(\xi) = \alpha e^{\beta \cdot R(\xi, o)}, w_\xi = \frac{p(\xi)}{p_R(\xi)} \tag{6}$$

where $p_R(\xi)$ is the probability of ξ inside the buffer, while w_ξ are the importance weights to balance the fact that ξ was sampled using $P_R(\Xi)$ rather than the current distribution $P(\Xi;\theta)$.

Importance sampling has two main advantages: it ensures that all the instances being sampled are possible since the rationality of impossible ones is 0. This prevents wasting computations on irrelevant cases. It also makes the sampling process progress more steadily towards instances with high rationality, since highly rational instances are sampled more often by using $P_R(\Xi)$ rather than $P(\Xi;\theta)$. This speeds up the convergence of the algorithm.

3.3 Training algorithm

Our proposed optimization procedure based on E-M with importance sampling is implemented by the following algorithm.

Algorithm 1 Rationality-Maximization

1: **procedure** RATIONALITY-MAXIMIZATION(o, k)
2: $\Delta\theta \leftarrow \infty$
3: **while** $\Delta\theta > k$ **do**
4: $\mathbf{\Xi} \sim P(\Xi; \theta_t)$ ▷ Sample a set of instances using θ_t
5: $R_{\mathbf{\Xi}} = $ COMPUTE-RATIONALITY($\mathbf{\Xi}, o$) ▷ Compute rationality
6: $P_R(\Xi).update(\mathbf{\Xi}, R_{\mathbf{\Xi}})$ ▷ Update the memory buffer
7: $\mathbf{\Xi_R} \sim P_R(\Xi)$ ▷ Sample from the memory buffer
8: $w_{\Xi_R} = \frac{p(\Xi_R;\theta_t)}{p_R(\Xi_R)}$ ▷ Compute the importance weights
9: $\Delta\theta \leftarrow R_{\Xi_R} \cdot w_{\Xi_R} \cdot \frac{d}{d\theta} p(\Xi; \theta_t)$ ▷ Compute $\Delta\theta$
10: $\theta_{t+1} \leftarrow \theta_t + lr \cdot \Delta\theta$ ▷ Update the parameters for the next iteration
11: **end while**
12: **end procedure**

Algorithm 1 performs the following steps: **Line 2-3:** The computation ends when no further progress can be made towards optimizing θ. **Line 4-6:** Randomly sample some planning instances using the current parameters at iteration t and compute their rationality (E-step). Store these in the memory buffer. **Line 7-8:** Sample from the memory buffer using importance sampling. **Line 9-10:** Update the model's parameters for the $t+1$ iteration using the instances sampled from the memory buffer.

4 Implementation in PDDL

We implement BDI models by specifying planning instances using the Planning Domain Description Language (PDDL). PDDL [13] is a standard language to specify planning domains for what is usually referred to as classical planning. A planning instance is obtained by specifying the tuple $\langle \mathcal{P}, \mathcal{A}, I, \mathcal{G}, \mathcal{O} \rangle$. Where I and $\mathcal{G}$ are the initial and goal state respectively, $\mathcal{O}$ is the set of objects available to ground the predicates $\mathcal{P}$, while $\mathcal{A}$ is the set of available actions to transition between states. The observer agent infers equivalent PDDL components $\xi = \langle \mathcal{P}^{act}_{obs}, \mathcal{A}^{act}_{obs}, I^{act}_{obs}, \mathcal{G}^{act}_{obs}, \mathcal{O}^{act}_{obs} \rangle$ that allow to compute intentions Π^{act}_{obs}. $\langle \mathcal{P}^{act}_{obs}, \mathcal{A}^{act}_{obs}, \mathcal{O}^{act}_{obs} \rangle$ is the inferred action schema $a \in A$, $\langle I^{act}_{obs} \rangle$ its inferred belief $b \in B$, while $\langle \mathcal{G}^{act}_{obs} \rangle$ the inferred desire $g \in G$. The probability distribution over the possible instances is defined as a combination of a Bernoulli distribution for the beliefs, and two categorical distributions for action schemas and goals.

$$
\begin{aligned}
P(\Xi, \theta) &= P(B; \theta_B) P(A; \theta_A) P(G; \theta_G) \\
P(B; \theta_B) &= \Pi_i P(p_i \in I^{act}_{obs}; \theta_{p_i}),\ p(p_i \in I^{act}_{obs}) = \theta_i \\
P(A; \theta_A) &= P(A | \{a_0, ..., a_n\}),\ p(A = \langle \mathcal{P}^{act}_{obs}, \mathcal{A}^{act}_{obs}, \mathcal{O}^{act}_{obs} \rangle | \{a_0, ..., a_n\}) = \theta_{ni} \\
P(G; \theta_G) &= P(G | \{g_0, ..., g_m\}),\ p(G = \langle \mathcal{G}^{act}_{obs} \rangle | \{g_0, ..., g_m\}) = \theta_{mi} \\
\sum_i \theta_{ni} &= 1, \sum_i \theta_{mi} = 1
\end{aligned}
$$

The rationality of a sequence of observations O in an PDDL instance ξ is measured as proposed in previous research [15]:

$$R(\xi, \pi_O) = \exp\{\tau(|\pi_{opt}| - |\pi_o|)\} \quad (7)$$

where $|\pi_{opt}|$ is the length of an optimal plan of ξ, while $|\pi_o|$ the length of the optimal plan constrained to contain O.

5 Experiments

We tested our model for a series of joint goal and belief recognition tasks, performed on an existing dataset for goal recognition in PDDL [14] on the following domains: *satellite, logistics, ferry, easy-ipc-gridworld, kitchen, intrusion-detection, campus*. For each domain we selected 10 random planning instances. For every planning instance beliefs and goals were randomized, while we kept the action schemas as fixed. Table 1 shows averages of several measures related to the original planning instances: number of operators, number of predicates, size of the initial state, size of the goal (number of predicates) and length of the optimal plans.

Domain	$\lvert\mathcal{A}\rvert$	$\lvert\mathcal{P}\rvert$	$\lvert I\rvert$	$\lvert G\rvert$	$\lvert\overline{\pi}\rvert$
intrusion	9.0	11.0	1.0	4.75	17.15
kitchen	29.0	23.0	2.0	1.0	10.6
satellite	5.0	12.0	62.8	6.8	16.55
campus	22.0	12.0	1.0	2.75	4.925
blocks-world	4.0	5.0	14.4	4.95	15.25
logistics	6.0	3.0	22.7	2.3	31.25
easy-ipc-grid	3.0	8.0	227.4	1.0	17.2
miconic	4.0	8.0	518.6	6.6	24.85
ferry	3.0	7.0	99.3	8.9	28.27

Table 1: Average instance measures over the tested planning domains. $|\mathcal{A}|$: number of operators, $|\mathcal{P}|$: number of predicates, $|I|$: size of the initial state, $|G|$: size of the goals, $|\overline{\pi}|$: length of the optimal plans.

For each tested sequence of observations we generated a randomized initial estimate of the actor's belief by using the original initial state of the problem, which in our case corresponds to the actor's true belief, further adding randomly generated ground predicates. The number of random ground predicates being added were equal to 10% of the number of possible ground predicates for that instance. The prior likelihood of every belief predicate, was set to 0.5, and the prior likelihoods over the goals were set to $\frac{1}{|G|} = 0.25$ (i.e. the observers started from a maximally entropic estimate of the actor). The memory was initialized with 200 randomly sampled valid planning instances (i.e. instances that reached their respective goal state from the initial state). Table 2 shows the measured hit rate on the correct goal by increasing percentages of observed actions, the columns are *%obs*: percentage of observations, *hit*: accuracy of recognized goals, $|I^+|$: beliefs size (number of predicates), R_{init}: rationality of the instances in memory priorly to training, R_{mem}: rationality of the instances in memory after training, R_{model}: avg. rationality of the instances from the model after training, D_{avg}: distance of the obtained intentions from the original observations, D_{min}: minimum distance of the obtained intentions from the original observations. D is a measure of state trajectory distance defined as:

$$D_{avg} = E_{\pi \sim P(\Pi|\xi,O),\xi \sim P(\Xi,\theta_R)}[\sum_{i \in 1..n} D(\pi_i, o_i)] \tag{8}$$

where $D(\pi_i, o_i)$ is the Jaccard distance between the i-th state obtained by unfolding plans coming from the learned $P(\Xi;\theta_R)$ and the i-th state computed using the ground-truth planning instance.

6 Discussion

Table 2 shows that we were able to jointly find with high accuracies, and for all the tested instances, the correct goal and belief behind the partial plans used as evidence. The rationality measures R_{init} of the original instances, and R_{model} for the final obtained instances, indicates that the algorithm correctly maximizes rationality. The small minimum pairwise state distances of intentions and observations show that some resulting intentions yield state transitions that are close to equal to the state transitions computed on the true instance, however, the larger D_{avg} indicates that the valid rational intentions are sampled from a broader belief space. In general, these measures suggest that the model correctly finds multiple rational interpretations in terms of goals and beliefs for a fixed sequence of observations, that are spread over a probabilistic space of beliefs, in a consistent way for all of the tested domains.

domain	%obs	hit	$\lvert I^+ \rvert$	R_{init}	R_{mem}	R_{model}	D_{avg}	D_{min}
logistics	0.30	1.00	63.69	0.21	0.56	0.05	0.40	0.08
logistics	0.50	1.00	63.69	0.21	0.86	0.23	0.34	0.14
logistics	0.70	1.00	63.69	0.21	0.93	0.36	0.31	0.17
blocks	0.30	0.80	38.30	0.46	0.87	0.21	0.49	0.18
blocks	0.50	0.80	38.30	0.46	0.91	0.24	0.45	0.09
blocks	0.70	0.80	38.30	0.46	0.98	0.34	0.42	0.13
grid	0.30	0.93	664.40	0.17	0.61	0.26	0.27	0.00
grid	0.50	0.93	664.40	0.17	0.74	0.24	0.24	0.02
grid	0.70	0.93	664.40	0.17	0.79	0.23	0.24	0.01
kitchen	0.30	0.80	11.60	0.47	0.87	0.38	0.68	0.21
kitchen	0.50	0.80	11.60	0.47	0.90	0.36	0.66	0.41
kitchen	0.70	0.80	11.60	0.47	0.91	0.41	0.63	0.38
campus	0.30	0.90	10.00	0.12	0.73	0.32	0.69	0.54
campus	0.50	1.00	10.00	0.12	0.82	0.24	0.66	0.41
campus	0.70	1.00	10.00	0.12	0.83	0.14	0.61	0.16
ferry	0.30	1.00	277.50	0.35	0.90	0.20	0.23	0.11
ferry	0.50	1.00	277.50	0.35	0.97	0.42	0.21	0.11
ferry	0.70	1.00	277.50	0.35	1.00	0.55	0.19	0.10
satellite	0.30	0.90	186.10	0.39	0.91	0.26	0.28	0.11
satellite	0.50	0.90	186.10	0.39	0.97	0.46	0.27	0.14
satellite	0.70	0.90	186.10	0.39	1.00	0.57	0.27	0.12
intrusion	0.30	1.00	10.60	0.74	0.89	0.37	0.82	0.64
intrusion	0.50	1.00	10.60	0.74	0.98	0.46	0.75	0.58
intrusion	0.70	1.00	10.60	0.74	1.00	0.56	0.70	0.56
miconic	0.30	1.00	1476.30	0.30	0.95	0.47	0.16	0.06
miconic	0.50	1.00	1476.30	0.30	1.00	0.59	0.15	0.05
miconic	0.70	1.00	1476.30	0.30	1.00	0.60	0.15	0.02
avg	0.30	0.93	304.28	0.36	0.81	0.28	0.45	0.21
	0.50	0.94	304.28	0.36	0.90	0.36	0.41	0.22
	0.70	0.94	304.28	0.36	0.94	0.42	0.39	0.18

Table 2: Average resulting measures for the tested domains. See text for additional details.

However, we had to employ a few tricks to contrast the complexity of computing probabilistic PDDL instances. In particular, populating the memory prior to training, and reusing results from a smaller number of observed plans was necessary to achieve high accuracy for longer sequences. In the absence of these two actions, the algorithm struggled to converge when long sequences of observations were provided. The reason for this is the difficulty in finding, from scratch, planning instances consistent with long plans.

The obtained accuracy is comparable with previous work on joint inference of belief and goal [3].

7 Conclusions

We have presented an algorithm for jointly inferring belief, goals, intentions and action schemas of a BDI agent by maximizing the rationality contained in the observations. The algorithm was implemented and evaluated on several standard PDDL domains. Our results demonstrate that a probability distribution for an actor's model can be constructed using prior assumptions about its action schemas and beliefs, combined with gathered observations as evidence. The intentional state of the agent (its committed plan of action) is a product of those. The proposed method was tested over several standard domains, where the actor's goal, beliefs and intentions were jointly inferred.

This work is related to many contributions in previous research, and attempts to better describe how to model an actor agent without assuming strong prior models. We showed how, building on just the assumption of rationality, it is possible to infer agents models in terms of their action schemas, beliefs, desires and intentions. We referred to the set of models induced by rationality as the equivalence class of rational models. We also proposed an algorithm to obtain such classes of models from observations. We implemented our method using PDDL and showed its applicability in multiple different domains.

Since the model space is usually very large, some starting assumptions on the agent model are necessary to make the proposed iterative procedure converge to a solution. This is expressed by the set of priorly known models of the actor. However, we make no assumption about these prior models used to construct the class of rational models. Intuitively, we expect that the richer they are in descriptive power, and the more similar they are to the observed agent, the better prediction capability they offer.

A relevant point that we would like to highlight is that the presented method based solely on rationality uses models and symbols that are internal to the observer, and therefore accessible for inference in autonomous robots that cannot directly access the state of other agents. This makes the model compatible with a first-order theory of mind setting. Despite its plausability in humans [6], and contrary to most of previous research, we do not use a model of how the actor perceives its environment, but focus only on the observations gathered by the observer. We however achieved accuracies comparable to methods explicitly modeling how the actor perceives. Future research could complement these methods for greater prediction accuracy.

References

1. Understanding and sharing intentions: The origins of cultural cognition. 28(5):675–691, 2005.

2. F. Alzetta, P. Giorgini, M. Marinoni, and D. Calvaresi. Rt-bdi: A real-time bdi model. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 16–29. Springer, 2020.
3. C. L. Baker and J. B. Tenenbaum. Modeling human plan recognition using bayesian theory of mind. *Plan, activity, and intent recognition: Theory and practice*, pages 177–204, 2014.
4. F. Bianco and D. Ognibene. Functional advantages of an adaptive theory of mind for robotics: a review of current architectures. In *2019 11th Computer Science and Electronic Engineering (CEEC)*, pages 139–143. IEEE, 2019.
5. T. Chakraborti, S. Sreedharan, Y. Zhang, and S. Kambhampati. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. *arXiv preprint arXiv:1701.08317*, 2017.
6. G. G. Cole and A. C. Millett. The closing of the theory of mind: A critique of perspective-taking. *Psychonomic bulletin & review*, 26(6):1787–1802, 2019.
7. S. Devin and R. Alami. An implemented theory of mind to improve human-robot shared plans execution. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 319–326. IEEE, 2016.
8. A. D. Dragan, K. C. Lee, and S. S. Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.
9. M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge. The belief-desire-intention model of agency. In *International workshop on agent theories, architectures, and languages*, pages 1–10. Springer, 1998.
10. T. Hellström and S. Bensch. Understandable robots: what, why, and how. *Paladyn-Journal of Behavioral Robotics*, 9(1):110–123, 2018.
11. A. Kulkarni, S. Srivastava, and S. Kambhampati. Signaling friends and head-faking enemies simultaneously: Balancing goal obfuscation and goal legibility. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1889–1891, 2020.
12. A. M. MacNally, N. Lipovetzky, M. Ramirez, and A. R. Pearce. Action selection for transparent planning. In *AAMAS*, pages 1327–1335, 2018.
13. D. McDermott. Pddl-the planning domain definition language. 1998.
14. R. Pereira, N. Oren, and F. Meneguzzi. Landmark-based heuristics for goal recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
15. M. Persiani and T. Hellström. Intent recognition from speech and plan recognition. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 212–223. Springer, 2020.
16. N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. A. Eslami, and M. Botvinick. Machine theory of mind. In *International conference on machine learning*, pages 4218–4227. PMLR, 2018.
17. R. Raileanu, E. Denton, A. Szlam, and R. Fergus. Modeling others using oneself in multi-agent reinforcement learning. In *International conference on machine learning*, pages 4257–4266. PMLR, 2018.
18. M. Ramírez and H. Geffner. Plan recognition as planning. In *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.
19. M. Shvo, T. Q. Klassen, S. Sohrabi, and S. A. McIlraith. Epistemic plan recognition. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1251–1259, 2020.