Progress Presentation

13.3. - 3.4.

Agenda

rational agent behaviour with I-POMDP

practical implementation with “Interactive Particle Filter”

Research question:

What kind of strategies do civilisations in the universe employ to ensure their survival, and how do these strategies change over time and space?

Interactive Partially Observable Markov Decision Process

Journal of Artificial Intelligence Research 24 (2005) 49-79 Submitted 09/04; published 07/05

A Framework for Sequential Planning in Multi-Agent Settings

Piotr J. Gmytrasiewicz PIOTR@CS.UIC.EDU
Prashant Doshi PDOSHI@CS.UIC.EDU
Department of Computer Science
University of Illinois at Chicago
851 S. Morgan St
Chicago, IL 60607

Abstract

This paper extends the framework of partially observable Markov decision processes (POMDPs) to multi-agent settings by incorporating the notion of agent models into the state space. Agents maintain beliefs over physical states of the environment and over models of other agents, and they use Bayesian updates to maintain their beliefs over time. The solutions map belief states to actions. Models of other agents may include their belief states and are related to agent types considered in games of incomplete information. We express the agents' autonomy by postulating that their models are not directly manipulable or observable by other agents. We show that important properties of POMDPs, such as convergence of value iteration, the rate of convergence, and piece-wise linearity and convexity of the value functions carry over to our framework. Our approach complements a more traditional approach to interactive settings which uses Nash equilibria as a solution paradigm. We seek to avoid some of the drawbacks of equilibria which may be non-unique and do not capture off-equilibrium behaviors. We do so at the cost of having to represent, process and continuously revise models of other agents. Since the agent's beliefs may be arbitrarily nested, the optimal solutions to decision making problems are only asymptotically computable. However, approximate belief updates and approximately optimal plans are computable. We illustrate our framework using a simple application domain, and we show examples of belief updates and value functions.

1. Introduction

We develop a framework for sequential rationality of autonomous agents interacting with other

Original I-POMDP paper

Journal of Artificial Intelligence Research 34 (2009) 297-337 Submitted 06/08; published 03/09

Monte Carlo Sampling Methods for Approximating Interactive POMDPs

Prashant Doshi PDOSHI@CS.UGA.EDU
Department of Computer Science
University of Georgia
415 Boyd GSRC
Athens, GA 30602

Piotr J. Gmytrasiewicz PIOTR@CS.UIC.EDU
Department of Computer Science
University of Illinois at Chicago
851 S. Morgan St
Chicago, IL 60607

Abstract

Partially observable Markov decision processes (POMDPs) provide a principled framework for sequential planning in uncertain single agent settings. An extension of POMDPs to multiagent settings, called interactive POMDPs (I-POMDPs), replaces POMDP belief spaces with interactive hierarchical belief systems which represent an agent's belief about the physical world, about beliefs of other agents, and about their beliefs about others' beliefs. This modification makes the difficulties of obtaining solutions due to complexity of the belief and policy spaces even more acute. We describe a general method for obtaining approximate solutions of I-POMDPs based on particle filtering (PF). We introduce the *interactive PF*, which descends the levels of the interactive belief hierarchies and samples and propagates beliefs at each level. The interactive PF is able to mitigate the belief space complexity, but it does not address the policy space complexity. To mitigate the policy space complexity – sometimes also called the curse of history – we utilize a complementary method based on sampling likely observations while building the look ahead reachability tree. While this approach does not completely address the curse of history, it beats back the curse's impact substantially. We provide experimental results and chart future work.

1. Introduction

Particle filtering method for approximating solutions

Finitely nested I-POMDPs

A finitely nested I-POMDP for agent i is given by

$$\text{I-POMDP}_{i,l} = (IS_{i,l}, A, T_i, \Omega_i, O_i, R_i)$$

where

$IS_{i,l}$ are level l interactive states. For $l = 1$,

$$IS_{i,1} = S \times \prod_{\substack{j=1 \\ j \neq i}}^{n} \Delta(S).$$

S are the possible environment/model states

$A = \prod_{j=1}^{n} A_j$ where A_j are the actions available to agent j. A restriction: for $a \in A$, only one a_j can be an actual action (others are always “no action”)

$T_i : S \times A \times S \to [0,1]$ is the transition function

Ω_i are possible observations of i

$O_i : S \times A \times \Omega_i \to [0,1]$ observation probabilities

$R_i : S \times A \to \mathbb{R}$ reward function

I-POMDP

Representing the model state

	Age	Visibility factor	Growth parameter 1	Growth parameter 2
Agent 1	10	0.5	0.6	20
Agent 2	53	1	0.9	30
Agent 3	0	1	0.4	40
Agent 4	12	0.9	0.5	12
Agent 5	5	0.1	0.3	20

$\in S$

I-POMDP

Beliefs about model state (level 0)

- Agents maintain beliefs $b_{i,l}(is)$ over the interactive beliefs $is \in IS_{i,l}$
- At level 0, beliefs are distributions over model states
- Interactive particle filter: a belief distribution is represented by a sample

$$\tilde{b}_{i,0} = \boxed{s_1} \quad \boxed{s_2} \quad \cdots \quad \boxed{s_k}$$

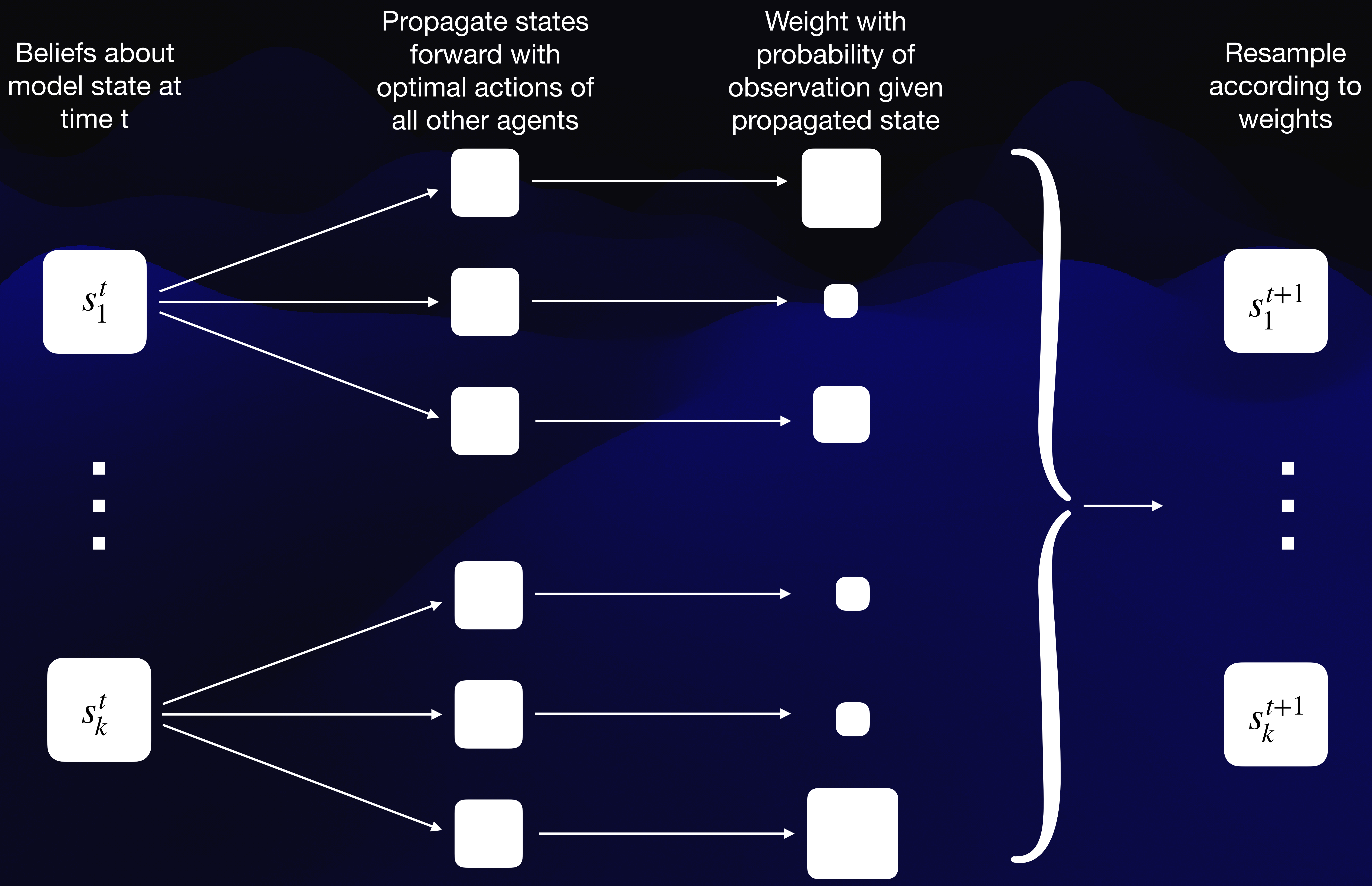
I-POMDP

Beliefs about others' beliefs (level 1)

$$\tilde{b}_{i,1} =$$

I-POMDP

Updating beliefs after acting and receiving an observation



I-POMDP

Optimal actions

Value of taking action a_i in belief state $b_{i,l}$ (Bellman equation):

$$U(b_{i,l}, a_i) = \text{reward this round} + \gamma \times \text{reward in future rounds}$$

$$= \int_{is \in IS_{i,l}} b_{i,l}(is) \sum_{a_{-i}} R_i(s, (a_i, a_{-i}))\mathbb{P}(a_{-i} \mid b_{i,l-1}) \,\mathrm{d}\, is$$

$$+ \gamma \int_{o_i \in \Omega_i} \mathbb{P}(o_i \mid b_{i,l}, a_i) U(SE(b_{i,l}, a_i, o_i)) \,\mathrm{d}\, o_i$$

Note: because in my model only one agent moves per turn, then if i gets to act, this sum will be over just ”no action” for every other agent

Update beliefs given own action and observation

I-POMDP

Optimal actions

Optimal action in belief state $b_{i,l}$ is therefore

$$\text{argmax}_{a_i} U(b_{i,l}, a_i)$$

I-POMDP

Optimal actions in practice

Value of taking action a_i in belief state $b_{i,l}$ (Bellman equation):

$U(b_{i,l}, a_i)$ = reward this round + γ × reward in future rounds

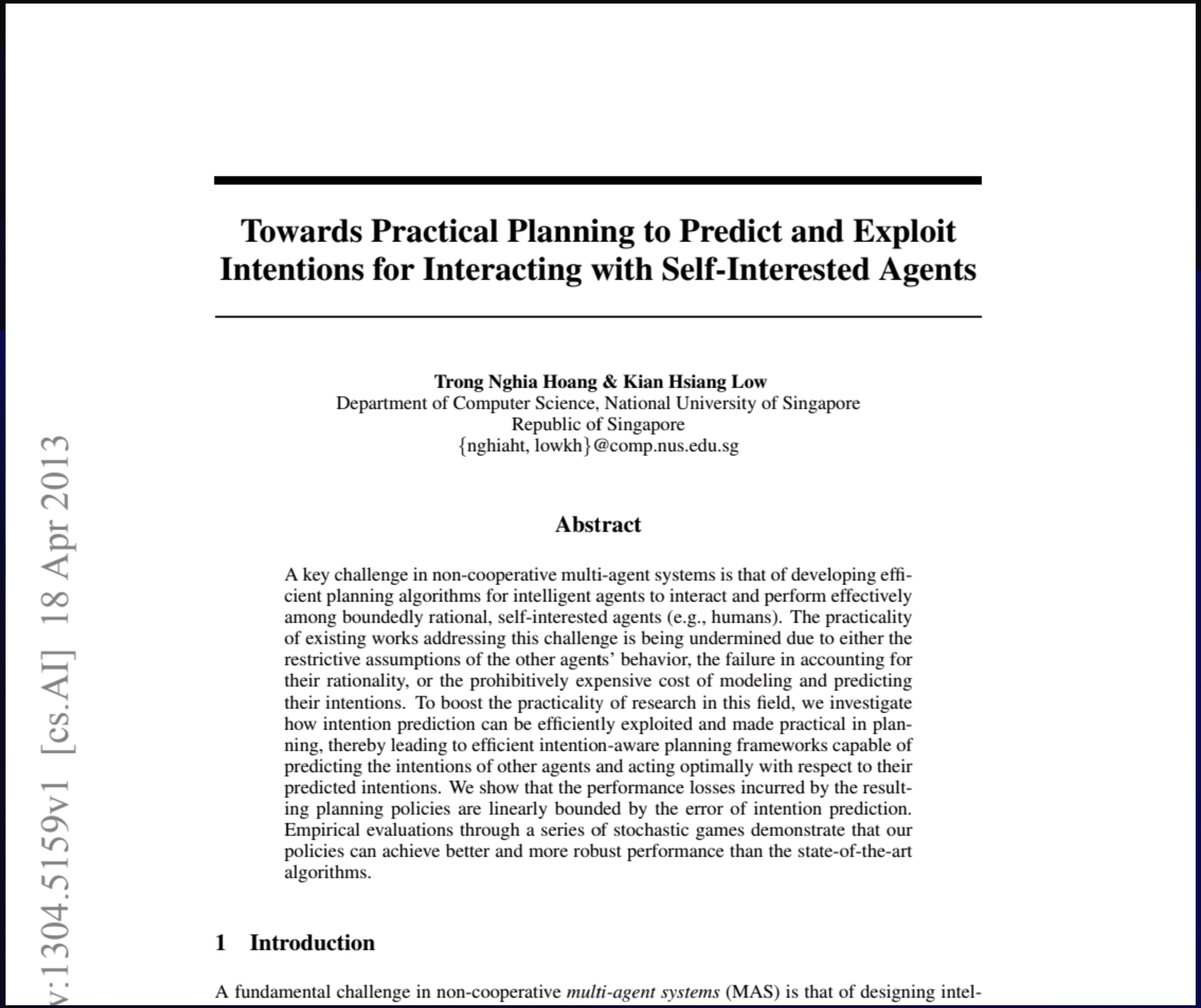
$$= \sum_{is \in \tilde{b}_{i,l}} \left(\sum_{\text{optimal } a_{-i}} R_i(s, (a_i, a_{-i})) + \gamma U(SE(b_{i,l}, a_i, o_i)) \right)$$

A single observation, sampled according to $O_i(s', a, \cdot\,)$ where s' is the forward propagated state, propagated by $a = (a_i, a_{-i})$

Problem

Slow.

I-POMDP Lite?



Towards Practical Planning to Predict and Exploit Intentions for Interacting with Self-Interested Agents

Trong Nghia Hoang & Kian Hsiang Low
Department of Computer Science, National University of Singapore
Republic of Singapore
{nghiaht, lowkh}@comp.nus.edu.sg

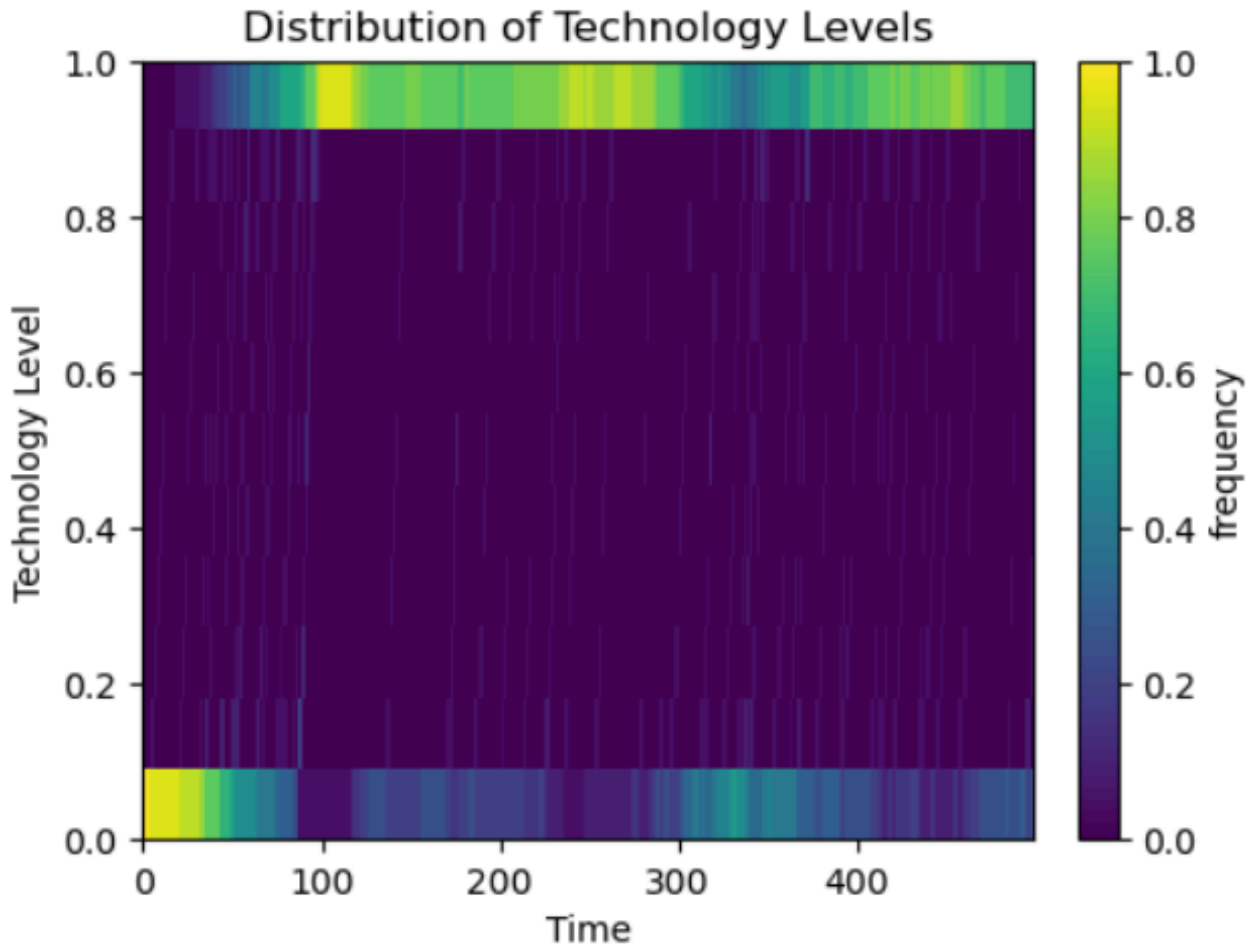
Abstract

A key challenge in non-cooperative multi-agent systems is that of developing efficient planning algorithms for intelligent agents to interact and perform effectively among boundedly rational, self-interested agents (e.g., humans). The practicality of existing works addressing this challenge is being undermined due to either the restrictive assumptions of the other agents' behavior, the failure in accounting for their rationality, or the prohibitively expensive cost of modeling and predicting their intentions. To boost the practicality of research in this field, we investigate how intention prediction can be efficiently exploited and made practical in planning, thereby leading to efficient intention-aware planning frameworks capable of predicting the intentions of other agents and acting optimally with respect to their predicted intentions. We show that the performance losses incurred by the resulting planning policies are linearly bounded by the error of intention prediction. Empirical evaluations through a series of stochastic games demonstrate that our policies can achieve better and more robust performance than the state-of-the-art algorithms.

1 Introduction

A fundamental challenge in non-cooperative *multi-agent systems* (MAS) is that of designing intel-

Technology belief distribution plot



num_steps: 500
num_agents: 20
decision_making: targeted
hostility_belief_prior: 0.1
speed_range: (0.3, 1)
takeoff_time_range: (10, 100)

Other issues

- Sometimes none of the sample model states are compatible with a given observation. This can happen when the number of samples is small.
  - An example of how this happens for a single sample:
    1. i is updating its level 1 beliefs.
    2. Therefore i propagates a single state in its beliefs forward. In the propagated state i thinks it should be able to observe civilisation j.
    3. But the observation does not include a techno signature from civilisation j, so i deems the propagated state impossible and gives it a weight of 0
  - solution: generate new beliefs that are by design compatible with the observation?

Other issues

- Mike still has not approved the research proposal. I will try to set up a meeting with him.