How to do Things with Words: A Bayesian Approach

Piotr Gmytrasiewicz

MULTIAGENT SYSTEMS GROUP

UIC UNIVERSITY OF ILLINOIS AT CHICAGO Department of Computer Science COLLEGE OF ENGINEERING

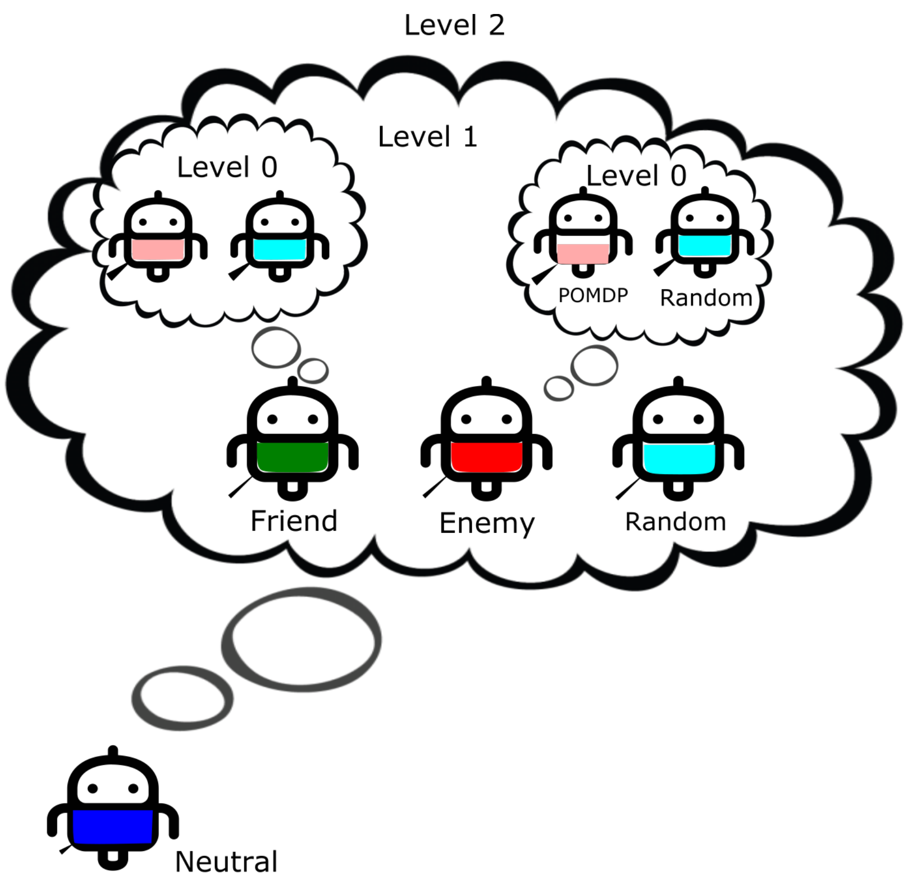
Life Science Exhibits
May 2021

Communication and Interaction among Agents

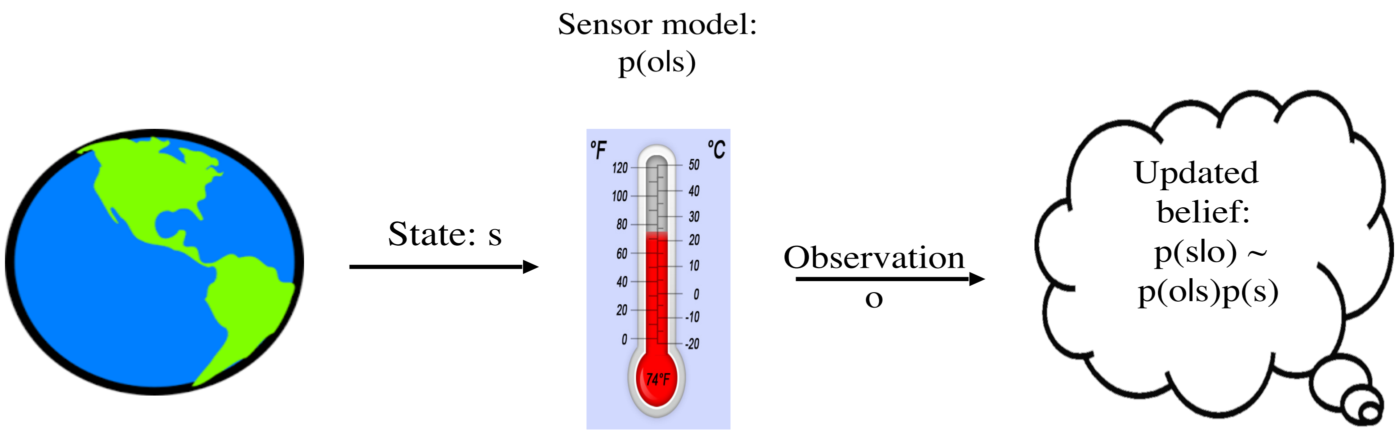
- **Why do ANYTHING?** Because it pays off! Imagine hunting elk with a companion, you see the animal but know you companion does not see it. Telling him/her about is ***beneficial***!
- Communication and interaction among self-interested agents in a partially observable and stochastic domain
- Application in military (Beautement, 2005), social robotics (Breazeal, 2008), cognitive science (Albert, 2018), economics (Evdokimov, 2019), etc.
- As more AI agents are becoming part of our social life, the study of emergent social behavior among communicating agents (both artificial and human) with **varied preferences** is vital

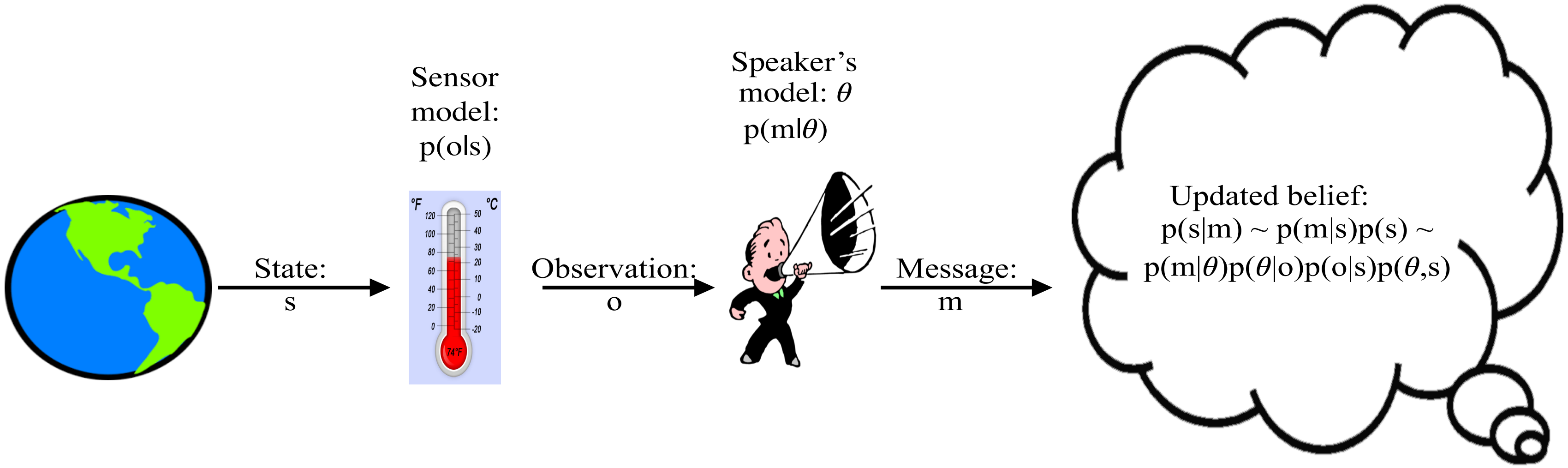
Deception and Theory of Mind

- The study of deception in a sequential decision-making scenario has rarely been explored
- As argued in (Isaac, 2017), AI needs to guard itself against malevolent humans and sometimes be able to deceive as well.
- To be able to cooperate, deceive or guard against deception, the agent needs to model the belief, intention, and preference of the other agent (Bridewell, 2011).
- A great deal of research in psychology establishes a connection of deception to the recursive theory of mind reasoning



Bayesian Belief Update





Background

Communicative Interactive Partially Observable Markov Decision Processes (CIPOMDPs) (Gmytrasiewicz, 2020)

$$CIPOMDP_i = \langle IS_{i,l}, A_i, \mathbb{M}, \Omega_i, T_i, O_i, R_i \rangle$$

$$IS_{i,l} = S \times M_{j,k}, l \geq 1,$$

$$M_{j,l} = \Theta_{j,l} \cup SM_j$$

$$\theta_{j,k} = \langle b_{j,k}, A_j, \Omega_j, T_j, O_j, R_j \rangle$$

$b_{j,k} \in \Delta(IS_{j,k})$ ***Interactive belief***

$$A = A_i \times A_j$$

$$T_i : S \times A \times S \rightarrow [0, 1]$$

$$O_i : S \times A \times \Omega_i \rightarrow [0, 1]$$

Belief Update in CIPOMDPs

$$b_i^t(is^t) = P(is^t | b_i^{t-1}, a_i^{t-1}, m_{i,s}^{t-1}, o_i^t, m_{i,r}^t) = \eta \sum_{is^{t-1}} b_i^{t-1}(is^{t-1}) \sum_{a_j^{t-1}} P(m_{j,s}^{t-1}, a_j^{t-1} | \theta_j^{t-1})$$

$$\times O_i(s^t, a^{t-1}, o_i^t) T_i(s^{t-1}, a^{t-1}, s^t) \sum_{o_j^t} \tau_{\theta_j^t}(b_j^{t-1}, a_j^{t-1}, m_{j,s}^{t-1}, o_j^t, m_{j,r}^t, b_j^t) O_j(s^t, a^{t-1}, o_j^t)$$

Planning in CIPOMDPs

$$U_i(\theta_i) = \max_{(m_{i,s}, a_i)} \Big\{ \sum_{is \in IS} b_{is}(s) ER_i(is, m_{i,s}, a_i)$$

$$+ \gamma \sum_{(m_{i,r}, o_i)} P(m_{i,r}, o_i | b_i, a_i) U_i(\langle SE_{\theta_i}(b_i, a_i, m_{i,s}, o_i, m_{i,r}), \hat{\theta}_i \rangle) \Big\}$$

Message Space

- Message space is defined as a set of discretized marginal probability distributions over a subset of variables in the interactive states of the agent
- The message space includes nil (silence), which is analogous to no-op operation in the physical action set.
- The message may provide the information about the interactive state, i.e., about physical state, other agents' models, their intentions, etc., marginalizing other variables of interactive state
- If message contains information about only the subset of possible interactive states, the remaining probability is uniformly distributed among the remaining variables (principle of indifference)

$$\forall_{s'} \in S'$$

$$m(s') = \frac{1 - \sum_{s \in S - S'} m(s)}{|S| - |S'|}$$

Communication for POMDP (0-level agent who does not model other agent(s)

Literal speaker

Generates a message reflecting its true belief about the physical states of the world b with probability 1 – α and all other messages including ’nil’ with probability α / (|M|–1)

The messages are then broadcasted to ”no one in particular” (NOIP), and do not take part in belief update for POMDP agent

Literal listener

can incorporate the incoming message as additional observation using augmented observation space and observation function

Augmented observation space and function for POMDP

Observation space now becomes a Cartesian product of usual observation space and message space

$$\Omega' = \Omega \times \mathbb{M}$$

The joint probability of observation and message received is obtained by combining the likelihood function for a physical observation with message distribution.

$$P(m_{i,r}|s) = \begin{cases} \frac{1}{|S|} & if m_{i,r} = nil \\ (m_{i,r}(s)) & otherwise \end{cases}$$

$$\forall_{m_{i,r} \in \mathbb{M}} \text{ and } \forall_{o \in \Omega} \text{ and } \forall_{s \in S}$$

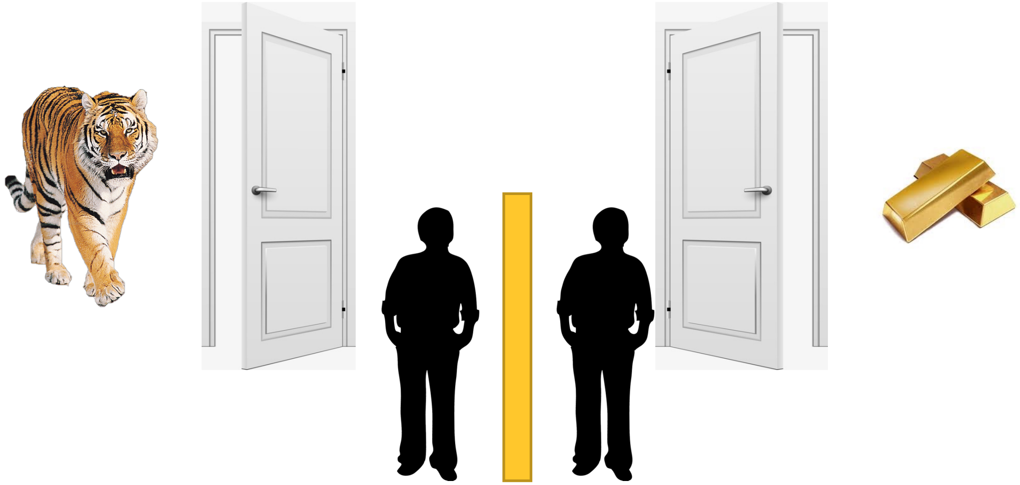
$$\begin{aligned} O'(s, a, o, m_{i,r}) &= \frac{P(o, m_{i,r}|s, a)}{\sum_{o, m_{i,r}} P(o, m_{i,r}|s, a)} \\ &= \frac{P(o|s, a) P(m_{i,r}|s, a)}{\sum_o P(o|s, a) \sum_{m_{i,r}} P(m_{i,r}|s)} \\ &= \frac{O(s, a, o) P(m_{i,r}|s)}{\sum_{m_{i,r}} P(m_{i,r}|s)} \end{aligned}$$

Algorithm - IPBVI-comm

- Exact value iteration quickly becomes intractable due to generation of large number of alpha-vectors which is exponential in observation space and message space
- Use Point based method to keep the size of the alpha-set in each iteration tractable
- The method only retains vectors which are optimal at the fixed set of belief points

Experiments and Results

Multi-agent tiger game



Action	State	GL	GR
<Listen, Listen>	TL	0.85	0.15
<Listen, Listen>	TR	0.15	0.85
<OpenRight, *>	*	0.5	0.5
<OpenLeft, *>	*	0.5	0.5
<*, OpenLeft>	*	0.5	0.5
<*, OpenRight>	*	0.5	0.5

Action / Transition	TL → TL	TL → TR	TR → TR	TR → TL
<OpenRight, *>	0.5	0.5	0.5	0.5
<OpenLeft, *>	0.5	0.5	0.5	0.5
<*, OpenLeft>	0.5	0.5	0.5	0.5
<*, OpenRight>	0.5	0.5	0.5	0.5
<Listen, Listen>	1.0	0.0	1.0	0.0

Conclusion

- Devised a technique to incorporate message into POMDP belief update
- Formalized the notion of sincerity and deception in terms of belief of the agent and messages in message space
- Adopted point-based solution method to CIPOMDPs to alleviate the complexities of considering communication as well as observations
- The analysis of computed policies shows the added sophistication of communication results in policies of superior quality, favoring the agents higher in cognitive hierarchy
- We showed how higher levels of theory of mind may allow an agent to disambiguate a sincere friend from deceitful foe based on received message and observation from the environment

Future Work

- We want to explore the higher depth of nesting, and more relaxed soft maximization criterion for action selection which can give rise to richer rational communicative behavior agents can engage in.
- Use of online planning method like Monte Carlo tree search for computing policy which provides scalability and with some variation, could accommodate continuous message space.
- Expanding message space to include other speech acts like question, assertion, intention, etc. to get richer agent communication language