Optimal Sequential Planning for Communicative Actions: A Bayesian Approach

Extended Abstract

Piotr Gmytrasiewicz
University of Illinois at Chicago
Chicago, IL
piotr@uic.edu

Sarit Adhikari
University of Illinois at Chicago
Chicago, IL
sadhik6@uic.edu

ABSTRACT

We build on the Interactive POMDP (IPOMDP) framework, which extends POMDPs to multi-agent settings, and include communication which can take place between the agents. While IPOMDPs endow the agents with models of their environments and models of other agents, we supplement IPOMDPs with communicative acts available to the agents to formulate Communicative IPOMDPs (CIPOMDPs). We treat communication as a type of action; hence decisions regarding communicative acts should be based on decision-theoretic planning using Bellman optimality principle, just as they are for all other actions. As in any form of planning, the results of actions need to be precisely specified. We use Bayes update to derive how agents update their beliefs in CIPOMDPs; updates are due to their actions, observations, messages they send to other agents, and messages they receive from others. Without communication CIPOMDPs reduce to IPOMDPs. Without other agents they all become classical POMDPs.

KEYWORDS

Single and multi-agent planning and scheduling; Speech act theory

ACM Reference Format:
Piotr Gmytrasiewicz and Sarit Adhikari. 2019. Optimal Sequential Planning for Communicative Actions: A Bayesian Approach. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

This paper reports on the Bayesian approach to specifying the consequences of communicative acts interwoven among physical actions and observations during an interaction with other agents and the physical world. Communicative actions are unlike physical acts since the function of communication is to change the agents' beliefs [9, 14, 17], not to change the physical environment. On the one hand communication is action [2] executed by a speaker, but on the other hand it is perception for the hearer.

Our contribution is to propose a principled approach to interaction and communication based on Bayesian decision theory and decision-theoretic planning. First, we build on interactive POMDPs [8] which allow agents to represent their state of knowledge about the physical states and about possible models of other agents. The ability of agents to model other agents has been called the theory of mind. Its usefulness while interacting with others has long been established in psychology, linguistics, economics and AI [6–8, 10, 13, 15, 16, 18]. We limit our attention to intentional models of other agents, which also are IPOMDPs, and in which agents' beliefs represent what they know about the state of the world and about other agents, including their preferences and beliefs about other agents' beliefs, others' beliefs about others and so on. Such beliefs are called interactive beliefs [1, 4, 8]. We augment IPOMDPs by allowing agents to send and receive messages, and call the resulting framework communicative IPOMDPs (CIPOMDPs). In finitely nested CIPOMDPs the models agents have of each other terminate at a finite level, called strategy level l, with "flat" POMDP models, like in IPOMDPs. There is no a priori bound on the value of the strategy level; agents are free to choose one as needed.

2 COMMUNICATIVE IPOMDPS

Communicative IPOMDPs (CIPOMDPs) build on IPOMDPs but include additional action of sending a message, m_s, and an additional observation - a message that could be received, m_r. Either message can be nil.

$$CIPOMDP_i = \langle IS_{i,l}, A, M, \Omega_i, T_i, O_i, R_i \rangle \quad (1)$$

where M is a set of messages the agents can send to and receive from each other (so that both m_s and m_r above are in M). All of the other elements are as defined previously for IPOMDPs except that the reward function has an additional argument; $R_i : S \times A \times M \rightarrow R$ so it can also depend on the messages i sends (messages can be costly.) The set of messages constitute the language of communication the agents share. We leave the exact specification of M for future work but we make an assumption that each message in M can be interpreted as a marginal probability distribution spanned on the agents' interactive state spaces IS_i (and IS_j). This allows agent i to send a message containing information about any variable(s) in i's belief space, and similarly for j. Messages with value nil (silence) contain no variables. The fact that the agents' beliefs and messages exchanged are probability distributions facilitates incorporation of information received into the agent's beliefs, subject to it's veracity. Note that while M is over the same state space as agent's beliefs we do not demand that it be in any way tied to the actual beliefs - agents are free to lie or be truthful in any way they find advantageous. To our knowledge our work is first not make the cooperative assumption.

Since interactive states encompass states of the world and other agents' nested models the message space defined above can contain

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

information about the world, information about other agents' anticipated actions, beliefs of other agents about the world and what they think others may do, and so on.

2.1 Belief Update

At any particular time step agents i and j perform physical actions and observe and also send and receive messages. Call the message i sent at time $t-1$ $m_{i,s}^{t-1}$, and one i received at time t $m_{i,r}^{t}$, and analogously for j. We assume all messages are in M and that message transmission is perfect. The belief update in CIPOMDPs has to update the probability of interactive state given the previous belief, action and observation, and given the message sent (at the previous time step) and received (at the current time): $P(is^t|b_i^{t-1}, a_i^{t-1}, m_{i,s}^{t-1}, o_i^t, m_{i,r}^t)$:

Proposition 1:

$$\begin{aligned} b_i^t(is^t) &= P(is^t|b_i^{t-1}, a_i^{t-1}, m_{i,s}^{t-1}, o_i^t, m_{i,r}^t) = \\ &= \alpha \sum_{is^{t-1}} b_i^{t-1}(is^{t-1}) \sum_{a_j^{t-1}} Pr(m_{j,s}^{t-1}, a_j^{t-1}|\theta_j^{t-1}) \times \\ &\times O_i(s^t, a^{t-1}, o_i^t) T_i(s^{t-1}, a^{t-1}, s^t) \\ &\times \sum_{o_j^t} \tau_{\theta_j^t}(b_j^{t-1}, a_j^{t-1}, m_{j,s}^{t-1}, o_j^t, m_{j,r}^t, b_j^t) O_j(s^t, a^{t-1}, o_j^t) \end{aligned} \tag{2}$$

Update in Proposition 1 is analogous to belief update in IPOMDPs when it comes to actions and observations. With respect to communication Proposition 1 combines three important elements. First, the updated belief depends on the agent's prior belief $b_i^{t-1}(is^{t-1})$, as should be expected. Second, the term $P(m_{j,s}^{t-1}, a_j^{t-1}|\theta_j^{t-1})$ quantifies the relation between the message i received from j and the model, θ_j, of agent j that generated the message.[1] This term is the measure of $j's$ sincerity, i.e., whether the message j sent reflects j's beliefs which are part of the model θ_j. We assume that agents are sincere to the extend that it pays off for them; we define this further below. Third, Proposition 1 includes the dependence of agent j's belief and the state of the world included in the interactive state $is =< s, \theta_j >$, both at time t and $t-1$; if j's observation function contained in θ_j is accurate i would expect that j's beliefs accurately reflect the state of the world s.

2.2 Decision-Theoretic Planning for Communication and Interaction

Given that belief update in CIPOMDPs[2] is analogous to belief update in IPOMDPs, we similarly proceed to define the Bellman equation which includes communicative actions and hard and soft criteria quantifying speaker's sincerity.

The belief update, defined in Proposition 1, over the whole space IS_i due to communication is again represented as a function SE so that the new belief is: $b_i^t = SE(b_i^{t-1}, a_i^{t-1}, m_{i,s}^{t-1}, o_i^t, m_{i,r}^t)$. $\tau_{\theta_i}(b_i^{t-1}, a_i^{t-1}, m_{i,s}^{t-1}, o_i^t, m_{i,r}^t, b_i^t)$ is defined as equal 1 when b_i^t is equal to $SE(b_i^{t-1}, a_i^{t-1}, m_{i,s}^{t-1}, o_i^t, m_{i,r}^t)$ and zero otherwise, and analogously for j (which is a factor in Proposition 1 above).

[1] Note that $m_{j,s}^{t-1} = m_{i,r}^t$ because message transmission is assumed to be perfect.

[2] Recall that either message may be nil – silence can be informative and takes part in the belief update.

The utility of interactive belief of agent i, contained in i's type θ_i, is:

$$\begin{aligned} U_i(\theta_i) = \max_{(m_{i,s}, a_i)} \Big\{ &\sum_{is \in IS} b_{is}(s) ER_i(is, m_{i,s}, a_i) + \\ &+\gamma \sum_{(m_{i,r}, o_i)} P(m_{i,r}, o_i|b_i, a_i) \times \\ &\times U_i(\langle SE_{\theta_i}(b_i, a_i, m_{i,s}, o_i, m_{i,r}), \hat{\theta}_i \rangle) \Big\} \end{aligned} \tag{3}$$

$ER_i(is, m_{i,s}, a_i)$ above is the immediate reward to i for sending $m_{i,s}$ and executing action a_i given the interactive state is and is equal to $\Sigma_{a_j} R_i(is, a_i, a_j, m_{i,s}) P(a_j|\theta_j)$, ($i$'s reward can depend on the cost of sending $m_{i,s}$ as we mentioned before.)

Equation above defines the utility of an interactive belief in θ_i and is the Bellman equation for interactive (physical and communicative) behavior. The agent's ability to compute optimal utility maximizing interactive behavior is the basis for rational interaction. The Bellman equation above describes the back-up operation during value-driven decision-theoretic search through an agent's interactive beliefs reachable by both agents' executing communicative and physical actions during interaction. It is a decision-theoretic version of multi-agent epistemic planning [3, 5, 11, 12].

An optimal message–action pair, $(m_{i,s}^*, a_i^*)$, agent i should execute (assuming infinite time horizon criterion with discounting) is an element of the set, $OPT(\theta_j)$, obtained by maximizing Eq. (3). This allows agents to predict which messages and actions are rational for other agents. As for IPOMDPs, there is a hard maximization criterion according to which i could model j as a strict optimizer and predict that j would only perform interactive actions in $OPT(\theta_j)$. Soft maximization defines the the probability of j sending $m_{j,s}$ and performing a_j as:

$$Pr(m_{j,s}, a_j|\theta_j) = \frac{exp[\lambda U_j(m_{j,s}, a_j)]}{\sum_{(m_{j,s}, a_j)} exp[\lambda U_j((m_{j,s}, a_j)]} \tag{4}$$

Equation (4) treats agents as rational and, when it comes to communication, is central to Rational Speech Acts model [9]. Equation above quantifies an agent's sincerity by tying the message it decides to send to it's beliefs (contained in θ_j) by modeling it as a self-interested rational speaker.

3 CONCLUSION AND FUTURE WORK

We presented an approach to decision-theoretic planning for communication and interaction by building on IPOMDPs. We added the capability for agents to exchange messages, derived Bayesian update of agents' beliefs due to message exchange, physical actions and observations. Further, we formulated Bellman optimality for interactive behavior. Our future work will include specification of the agent communication language and principled investigation of deceptive communicative behavior. We believe that agents can protect themselves from being lied to by out-thinking the other agent in terms of depth of the nested theories of mind and in terms of the time horizon.

REFERENCES

[1] Robert J. Aumann. 1999. Interactive Epistemology I: Knowledge. *International Journal of Game Theory* 28 (1999), 263–300.

[2] J. L. Austin. 1962. *How to do Things with Words.* Clarendon Press.
[3] Thomas Bolander and Mikkel Birkegaard Andersen. 2011. Epistemic Planning for Single- and Multi-Agent. *Journal of Applied Non-Classical Logics* 21 (2011), 9–34.
[4] Prashant Doshi and Piotr Gmytrasiewicz. 2009. Monte Carlo Sampling Methods for Approximating Interactive POMDPs. *Journal of AI Research* 34 (2009), 297–337.
[5] Thorsten Engesser, Thomas Bolander, Robert MattmÃijller, and Bernhard Nebel. 2017. Cooperative Epistemic Multi-Agent Planning for Implicit Coordination. *Electronic Proceedings in Theoretical Computer Science (ISSN: 2075-2180) (DOI: http://dx.doi.org/10.4204/EPTCS.243.6)* 243 (2017).
[6] Chris Frith and Uta Frith. 2005. Theory of mind. *Current Biology* 15, 17 (2005), R644 – R645. https://doi.org/10.1016/j.cub.2005.08.041
[7] Vittorio Gallese and Alvin Goldman. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2, 12 (1998), 493 – 501. https://doi.org/10.1016/S1364-6613(98)01262-5
[8] Piotr Gmytrasiewicz and Prashant Doshi. 2005. A Framework for Sequential Planning in Multiagent Settings. *Journal of Artificial Intelligence Research* 24 (2005), 49–79. http://jair.org/contents/v24.html.
[9] Noah D. Goodman and Daniel Lassiter. 2014. Probabilistic Semantics and Pragmatics: Uncertainty in Language and Thought, A draft chapter. In *Wiley-Blackwell Handbook of Contemporary Semantics âĂŤ second edition, https://web.stanford.edu/ ngoodman/papers/Goodman-HCS-final.pdf*, Shalom Lappin and Chris Fox (Eds.).
[10] H. P. Grice. 1975. Logic and Conversation. In *Studies in Syntax and Semantics III: Speech Acts*, P. Cole and J. Morgan (Eds.). Academic Press, 41–58.
[11] W. V. D. Hoek and M. Wooldridge. 2002. Tractable Multiagent Planning for Epistemic Goals. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002)*. 1167âĂŞ1174.
[12] Filippos Kominis and Hector Geffner. 2015. Beliefs in Multiagent Planning: From One Agent to Many. In *Proceedings of ICAPS*. 157–155.
[13] Alan M. Leslie, Ori Friedman, and Tim P. German. 2004. Core mechanisms in âĂŸtheory of mindâĂŹ. *Trends in Cognitive Sciences* 8, 12 (2004), 528 – 533. https://doi.org/10.1016/j.tics.2004.10.001
[14] David Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8, 1 (1979), 339–359.
[15] Ahti Pietarinen. 2007. *Game Theory and Linguistic Meaning*. Vol. 18. Elsevier: Current in the Semantics/Pragmatics Interface.
[16] Karen Shanton and Alvin Goldman. 2010. Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 4 (2010), 527–538. https://doi.org/10.1002/wcs.33
[17] R. Stalnaker. 1978. Assertion. In *Syntax and Semantics 9: Pragmatics*, P. Cole (Ed.). Academic Press.
[18] Adam Vogel, Christopher Potts, and Dan Jurafsky. 2013. Implicatures and nested beliefs in approximate decentralized-pomdps. ACL (2013).