Hierarchical Reinforcement Learning in Communication-Mediated Multiagent Coordination[*]

Felix Fischer Michael Rovatsos Gerhard Weiss
Department of Informatics
Technical University of Munich
85748 Garching, Germany
{fischerf,rovatsos,weissg}@cs.tum.edu

1. Introduction

Over the past decade, reinforcement learning (RL; e.g., see [9]) has been an active area of AI research in general and of research on agents and multiagent systems (MASs) in particular. In the original *Markov decision process* (MDP; e.g., see [5]) formulation of RL, other agents an agent is coexisting and interacting with are treated as part of its environment. The inability of MDPs to model multiple adaptive agents has explicitly been identified as the main drawback of this approach [4]. As a consequence, interest has grown in extending the RL framework to explicitly take into account other agents as autonomous and self-interested entities (see [8] for an overview).

In this paper, we follow this line of research while focussing on *communication-mediated* multiagent coordination problems. The idea here is that "physical" acting can be preceded by communication to allow for a prediction of actions to come. By assuming that this kind of communication does not manipulate the environment (i.e. hardly affects the states agents find themselves in) and does not have effects w.r.t. utility, we can view the exchanged messages as symbols that "encode" anticipated courses of physical action. This is in accordance with the model of communication we have laid out in [6].

We make two contributions to the solution of communication-mediated multiagent coordination problems:

1. We apply *hierarchical* RL methods [1] to the problem of communication learning.
2. We suggest powerful policy abstractions (so-called *interaction frames*) that enable generalisation over communication strategies expressed in speech-act-based agent communication languages.

2. Interaction frames

Interaction frames are a key concept of the abstract social reasoning architecture InFFrA proposed in [7]. There, they describe patterns of interaction that can be used strategically by knowledge-based agents in a reasoning process called *framing* to guide their communicative behaviour.

For the scope of this paper, it suffices to look at (interaction) frames as *policy abstractions* (in the sense of MDP policies). This interpretation forms the basis of a formal model of InFFrA called m^2InFFrA (where the m^2 stands for "Markov-square" and hints at the underlying hierarchical two-level MDP view), details of which can be found in [3].

In m^2InFFrA, a frame describes a set of two-party, discrete, turn-taking interaction *encounters* which can be thought of as conversations between two agents. A sequence of message patterns (i.e. messages containing variables) called *trajectory* specifies the surface structure of the encounters described by the frame, while a list of *variable substitutions* captures the values of variables in the trajectory in previously experienced interactions. Each substitution also corresponds to a set of logical *conditions* that were required for and/or precipitated by execution of the trajectory in the respective encounter. Finally, *trajectory occurrence* and *substitution occurrence* counters record the frequency with which the frame has occurred in the past.

3. Frames and options

In terms of MDPs, an m^2InFFrA frame can be seen as a *policy abstraction* describing a manageable "chunk" of a communication process that can be invoked as MDP decision and then executed until certain conditions apply (typically, until it has been executed completely, until a message is uttered not matching its trajectory, until its context conditions are no longer satisfied, or until further execution is considered undesirable according to some heuristics).

[*] A full version of this paper containing further technical details and experimental results can be retrieved from `http://www7.in.tum.de/~rovatsos/papers/frw_aamas04.ps`

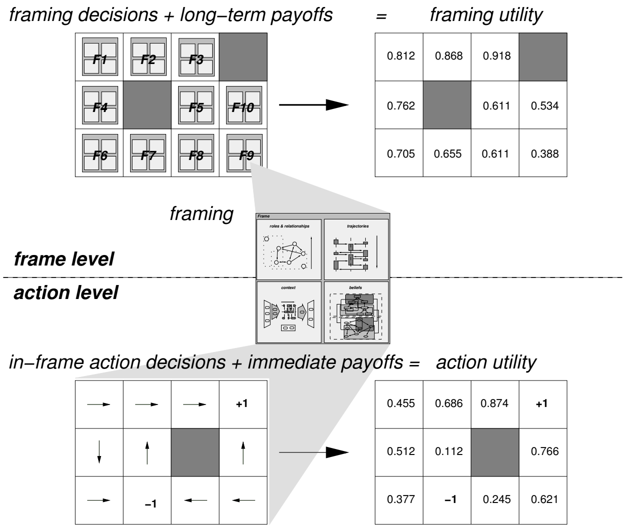

Figure 1. Frame-based hierarchical view of communication-mediated MDPs.

This interpretation allows us to combine the principles of InFFrA with the hierachical RL framework of *options* [10], which is based on augmenting the sets of admissible "primitive" actions by sets of so-called options, consisting of an input set of states in which the option is admissible, a (stationary, stochastic) policy that is followed when the option is invoked, and a (stochastic) termination condition.

While due to space limitations we cannot go into the technical details of how hiererachical RL and interaction frames are combined, what is important is the social reasoning and learning view that we obtain by using this approach, namely a two-level MDP as depicted in figure 1:

- At the *frame level*, the agent chooses a frame as a communication policy that may be used over an extended period of time, depending on whether it can be completed successfully. We employ SMDP Q-learning [2] to learn long-term "framing" behaviour and derive optimal strategies for frame selection from experience. Abstract representations of the goal of a conversation are used as the states of this upper-level MDP.

- At the *action level*, we have to determine which concrete instance of a frame to select so as to optimise the outcome of a conversation. Remembering that frames contain message patterns that may allow for additional choices (e.g. regarding which argument to use in an argumentation dialogue), we use *adversarial* search to maximise expected utility considering the other's potential reactions.

4. Conclusions

This paper proposed an approach for combing hierarchical RL with interaction frames as knowledge-level communication patterns that can be viewed as policy abstractions. In addition to the complexity-reducing aspects of hierarchical RL, this constitutes an important step to improve methods that aid in the construction of agents able to reason about communication patterns. At the same time, our research bridges the gap between machine learning techniques on one side and the design of communication languages and interaction protocols for knowledge-based agents on the other.

References

[1] A. G. Barto and S. Mahadevan. Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, 13(4):41–77, 2003.

[2] S. J. Bradtke and M. O. Duff. Reinforcement learning methods for continuous-time Markov decision problems. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 393–400. The MIT Press, 1995.

[3] F. Fischer. Frame-based learning and generalisation for multiagent communication. Diploma Thesis. Department of Informatics, Technical University of Munich, 2003.

[4] M. L. Littman. Markov games as a framework for multiagent reinforcement learning. In *Proceedings of the 11th International Conference on Machine Learning (ML-94)*, pages 157–163, New Brunswick, NJ, 1994. Morgan Kaufmann.

[5] M. L. Puterman. *Markov Decision Processes*. John Wiley & Sons, New York, NY, 1994.

[6] M. Rovatsos, M. Nickles, and G. Weiß. Interaction is Meaning: A New Model for Communication in Open Systems. In J. S. Rosenschein, T. Sandholm, M. Wooldridge, and M. Yokoo, editors, *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*, Melbourne, Australia, 2003.

[7] M. Rovatsos, G. Weiß, and M. Wolf. An Approach to the Analysis and Design of Multiagent Systems based on Interaction Frames. In M. Gini, T. Ishida, C. Castelfranchi, and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, Bologna, Italy, 2002.

[8] Y. Shoham, R. Powers, and T. Grenager. Multi-agent reinforcement learning: a critical survey. Technical report, Stanford University, 2003.

[9] R. S. Sutton and A. G. Barto. *Reinforcement Learning. An Introduction.* MIT Press/A Bradford Book, Cambridge, MA, 1998.

[10] R. S. Sutton, D. Precup, and S. P. Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1-2):181–211, 1999.