

# Learning with whom to communicate using Relational Reinforcement Learning

## (Extended Abstract)

Marc Ponsen  
University of Maastricht, The Netherlands  
m.ponsen@micc.unimaas.nl

Tom Croonenborghs  
Department IBW  
KH Kempen, Belgium  
tom.croonenborghs@khk.be

Karl Tuyls  
Eindhoven University of Technology, The Netherlands  
k.p.tuyls@tue.nl

Jan Ramon  
Declarative Languages and Artificial Intelligence Group  
K.U.Leuven, Belgium  
jan.ramon@cs.kuleuven.be

Kurt Driessens  
Declarative Languages and Artificial Intelligence Group  
K.U.Leuven, Belgium  
kurt.driessens@cs.kuleuven.be

### Categories and Subject Descriptors

I.2.6 [Computing methodologies]: Artificial intelligence—*Learning*

### Keywords

Reinforcement Learning, Multi-Agent Systems, Relational Reinforcement Learning

## 1. INTRODUCTION

Relational reinforcement learning (RRL) has emerged in the machine learning community as a new promising subfield of reinforcement learning (RL) (e.g. [1]). It upgrades RL techniques by using relational representations for states, actions and learned value-functions or policies to allow more natural representations and abstractions of complex tasks. This leads to a serious state space reduction, allowing to better generalize and infer new knowledge.

So far, most of the work in RRL has focused on single agent situations, i.e., environments in which only one agent operates and learns. Because multi-agent problems contain, besides a complex environment, a dynamic intermingle of relations between the agents, it is reasonable to assume this new paradigm can also contribute to the Multi-agent Systems (MAS) domain. Different approaches have been introduced to solve this multi-agent learning problem ranging from joint action learners to individual local Q-learners. They however all use a state space which is in propositional form, where by no means relations between different features or agents are expressed or exploited. Moreover, no method has tried to model the dynamic intermingle of relations between agents to simplify the learning of specific tasks. An important drawback on using a propositional representation is the difficulty to represent states that are defined by the objects present in the state and the relationships between these objects. Moreover, agents in a MAS can be viewed as objects with varying properties that also share a number of relations between them. To apply RL in such complex environments, a structural or relational representation is needed.

**Cite as:** Learning with Whom to Communicate using Relational Reinforcement Learning, (Extended Abstract), Marc Ponsen, Tom Croonenborghs, Karl Tuyls, Jan Ramon, Kurt Driessens, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 1221–1222

Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org), All rights reserved.

In this paper, we will focus on learning a communication policy for agents in an abstract coordination task.

## 2. RELATIONAL MULTI-AGENT RL

The complexity of designing a MAS and even more when trying to learn behaviors of agents within a MAS, stems from three important sources. First, agents in a MAS can interfere with each other's actions, plans or even goals. Hence, agents need to take these external influences into account and need to learn which other agents to keep in mind when optimizing ones own behavior. We believe that relational representations can be of great help when trying to define and learn these spheres of influence in MAS. Because of the inherent abstraction that must be used with relational representations, the agents can themselves decide which region of the state space to focus on. A second source of complexity is partial observability. Using multiple agents instead of a single, all-encompassing agent makes more sense if the MAS system does not require all agents to have access to all details of their combined environment. We are convinced that relational representations are beneficial when dealing with partial observability because they allow the agent to abstract away details of the environment which are unimportant to the decision making process and focus their attention on the state-defining structures they do want to base decisions on.

A third cornerstone, one that is linked to both interference between agents and partial observability is the need for or opportunity for communication between agents. Communication can be used to limit the influence of partial observability or to avoid unwanted interference between agents. However, in MAS communication is often assumed to be either unreliable or costly. Therefore, we consider the communication strategy of agents a good starting point to introduce relational reinforcement learning into MAS.

## 3. EMPIRICAL EVALUATION

We consider a population of agents that each have to learn how to solve a number of tasks. More specifically, we use a MAS containing 10 agents and 10 different tasks. Every agent is assigned to a random task in the beginning of an episode and receives a reward of 1 if he solves his task and neither reward nor punishment otherwise. An episode ends when every agent executed 100 actions or solved his task. The number of possible actions (henceforth called *primitive actions*) an agent can execute during a certain task can

vary from task to task. To solve a task the agent has to execute the *optimal* action, which is a single action from the possible primitive actions for that task. The tasks all have a different difficulty: In the first task there are only 5 primitive actions available, for the second task this is doubled to 10 primitive actions and so forth until 2560 actions for the tenth task. Some agents have a priori knowledge about certain tasks, these agents are called *experts* for that task. In our test environment, agents and tasks have a number of properties and we define an agent as an expert for a certain task if he shares a property with that task. The experiment is set up in such a way that every agent is expert for exactly one different task and hence there is exactly one expert for every task.

Agents have the possibility to communicate with each other. Agents can seek advice from other agents by asking them which action they would execute according to their current policy in the current task of the inquirer. Therefore, the action space of an agent for some task consists of the number of primitive actions for this task and the “*advice*”-actions, one for every agent in the environment.

To study the integration of MAS and RRL, we should take care not to make the learning task too complex at once, as a mix up of the many different effects could make results difficult to interpret. Therefore, we limit the influences of agent interference and partial observability in this paper by letting the agents all work on their own task without the possibility to interfere with each other.

As RRL technique we employ relational  $Q$ -learning where an incremental relational tree learner RRL-TG [2] is used to learn a generalized  $Q$ -function. To guide this tree building a declarative bias is specified that contains a language bias to specify the predicates that can be used as tests in the tree to partition the state and action space. For all agents this language bias contains predicates that identify the task and the action taken by the agent. Furthermore, it contains predicates that check certain properties of the task at hand and properties of the other agents.

We will first discuss the case where the experts are *pure* experts, i.e. the agent only has access to the task’s optimal action, and therefore has no need for exploring the action space for this particular task. Figure 2 clearly shows the gain that agents with the capability of communication have over agents without this capability (in the case of pure experts). Moreover, the graph shows two different results for ‘pure experts’. For the relational one, the agents use the language bias as described earlier. The propositional one uses a restricted language bias to simulate a propositional tree learner. It can be concluded that the generalization of relational learning is essential to take optimal advantage of the added communication. The policy learned by the agents in the relational setting is shown in Figure 1. The tree very concisely illustrates the optimal relational policy, namely one should seek advice from a pure expert. Otherwise if no expert exists for the task, the agent itself is the expert.

Next, we investigate the influence of the prior knowledge that agents have. To investigate this, we do not only consider pure experts but introduce different levels of expertise: agents may have some knowledge about the task at hand, i.e. they know that the optimal action is in some subset of all primitive actions for that task and consequently only need to explore a constrained part of the action space. In the setting ‘no experts’, the agents have no prior knowledge, i.e. the possible actions for the agents are all the primitive actions and advice actions. In the setting ‘optimal + advice’, the experts can only execute the optimal action as primitive action, but they can also ask for advice and hence they still have to learn to select the primitive action instead of asking for advice. We also investigated two intermediate settings where a subset of all primitive actions is selected, namely 50% resp. 75% of the possible prim-

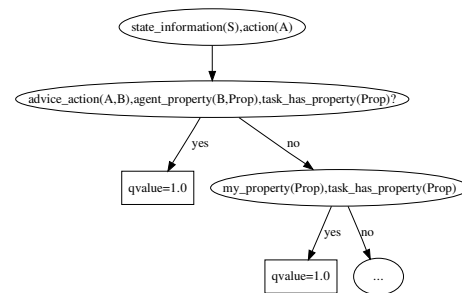


Figure 1: Policy with communication (relational)

itive actions. One can clearly conclude from the graph that when using relational learning communication aids the agents even when prior knowledge is limited. In the early stages of learning, the extra possibility of seeking advice does not hurt the agent, only in the very beginning the performance is slightly worse due to the bigger action space. On the other hand, once some agents learned to solve some tasks, this knowledge is distributed through the population resulting in a performance gain.

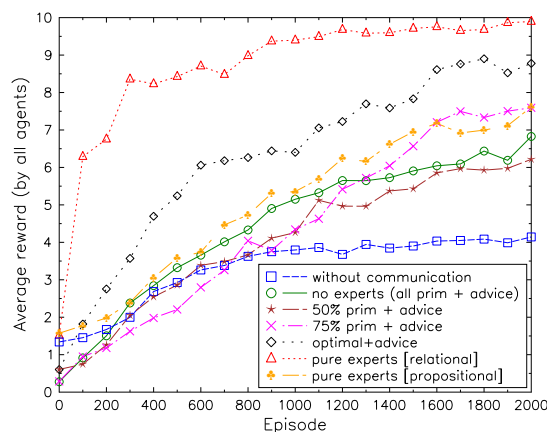


Figure 2: Rewards obtained by freezing the  $Q$ -function approximation and following a greedy policy to test for optimality (averaged over 5 runs)

#### 4. ACKNOWLEDGEMENTS

Marc Ponsen is sponsored by the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024. Jan Ramon and Kurt Driessens are post-doctoral fellows of the Fund for Scientific Research (FWO) of Flanders.

#### 5. REFERENCES

- [1] K. Driessens. *Relational Reinforcement Learning*. PhD thesis, Department of Computer Science, Katholieke Universiteit Leuven, 2004.
- [2] K. Driessens, J. Ramon, and H. Blockeel. Speeding up relational reinforcement learning through the use of an incremental first order decision tree learner. In *Proceedings of the 12th European Conference on Machine Learning*, volume 2167 of *LNAI*, pages 97–108. SV, 2001.