Context-sensitive reward shaping for sparse interaction MAS\textsuperscript{1}

Yann-Michaël De Hauwere \textsuperscript{a} Sam Devlin\textsuperscript{b} Daniel Kudenko \textsuperscript{b}
Ann Nowé \textsuperscript{a}

\textsuperscript{a} Computational Modeling Lab, Vrije Universiteit Brussel
{ydehauwe,anowe}@vub.ac.be
\textsuperscript{b} Department of Computer Science, University of York
{sam.devlin,daniel.kudenko}@york.ac.uk

Abstract
This paper describes the use of context aware potential functions in a multi-agent system in which the interactions between agents are sparse to guide agents towards the desired solutions. These sparse interactions mean that the interactions between the agents only occur sporadically, unknown to the agents a priori, in certain regions of the state space. During these interactions, agents need to coordinate in order to reach the global optimal solution.

We demonstrate how different reward shaping functions can be used on top of Future Coordinating Q-learning (FCQ); an algorithm capable of automatically detecting when agents should take each other into consideration. Using FCQ-learning, coordination problems can even be anticipated before the actual problems occur, allowing the problems to be solved timely. We evaluate our approach on a range of gridworld problems, as well as a simulation of Air Traffic Control.

1 Introduction

In many environments, agents are trying to achieve different subgoals based on the context the agent is currently in. For example, this context could be the high level location of the agent (the sector it is patrolling or the room it is currently in), where the subgoal is the next victim that has to be rescued after a disaster, or the next flag that needs to be collected in a flag domain. In multi-agent systems (MAS) this context might also be defined by the interactions that occur between the agents. Many MAS are characterised by the fact that agents only influence each other in particular regions of the state space, for instance consider autonomous guided vehicles in a warehouse. These vehicles mostly influence each other around the entrances of corridors in the main hallway. In situations where agents are not influencing each other, these agents should also not been taken into consideration and single agent RL can be applied. Most research around these sparse interactions focusses on learning when agents should coordinate their actions \cite{4} or learning when agents should augment their state space to include information from other agents \cite{2, 5}. This is achieved by using the reward signal to detect when coordination is beneficial or when agents can safely be ignored. In each of these contexts (i.e. acting individually or required to coordinate with another agent), the agent is often trying to reach a different subgoal (i.e. reach its individual goals vs solve the coordination problem with the other agent) and hence could benefit from using a different, more appropriate shaping function.

In this paper, the context of an agent is defined through the sparse interactions it has with other agents. Moreover, the influence of these interactions will only be reflected several time steps in the future, i.e.

\textsuperscript{1}Full paper published accepted for publication in 04/2013 in the Knowledge Engineering Review Journal
the reward signal is delayed. We build upon Future Coordinating Q-learning (FCQ-learning) [1], which is capable of detecting these future coordination problems and extend it by means of context-sensitive reward shaping; a novel design of potential functions for reward shaping.

2 Context-sensitive reward shaping

An agent, learning in an environment, can have different subgoals it is trying to achieve while accomplishing the global goal of the task at hand. Incorporating all these subgoals into one shaping function is not feasible. The idea we present here is to have different appropriate shaping functions, that depend on the context the agent is currently in. This context is defined by the subgoal it is currently trying to accomplish. The shaping function for a context will guide the agent towards achieving the particular goal of that context. These shaping functions can be defined for a single agent, or can be generated from a joint plan including the other agents’ state. As acknowledged by [3], individual plans might contain conflicting knowledge, which results in agents interfering with each other. On the other hand, shaping functions based on joint plans require prior coordination and knowledge of interaction states. As this may not always be possible, we propose, to use a different shaping function when acting individually or when coordinating. If these contexts can be detected automatically, agents can autonomously switch to a different shaping function for the particular context the agent is currently in. This allows the designer of the system to have multiple simple shaping functions, rather than try to build one shaping function that covers all the subtleties that occur when multiple agents act in the same environment.

We implement our approach to context-sensitive reward shaping by using FCQ-learning to detect the different contexts. FCQ-learning samples the state space and will automatically augment certain states to include state information about other agents if the agent is influenced by them. This means that an agent can be in one of following contexts:

1. **Individual**: The agent is not influenced by any other agent and only uses local state information to select its actions.
2. **Coordinating**: The agent is influenced by another agent and acts using augmented state information.

3 Conclusion

In the full paper we have demonstrated empirically how, in multi-agent systems, different shaping functions can be used, depending on the context of the agent, i.e. is it interacting with another agent or is it not. This distinction allows us to benefit from both the speedup achieved by FCQ-learning and the speedup obtained through reward shaping.

References


