Chapter 13
Effects of Shaping a Reward on Multiagent Reinforcement Learning

Sachiyo Arai
Chiba University, Japan

ABSTRACT

The multiagent reinforcement learning approach is now widely applied to cause agents to behave rationally in a multiagent system. However, due to the complex interactions in a multiagent domain, it is difficult to decide the each agent’s fair share of the reward for contributing to the goal achievement. This chapter reviews a reward shaping problem that defines when and what amount of reward should be given to agents. We employ keepaway soccer as a typical multiagent continuing task that requires skilled collaboration between the agents. Shaping the reward structure for this domain is difficult for the following reasons: i) a continuing task such as keepaway soccer has no explicit goal, and so it is hard to determine when a reward should be given to the agents, ii) in such a multiagent cooperative task, it is difficult to fairly share the reward for each agent’s contribution. Through experiments, we found that reward shaping has a major effect on an agent’s behavior.

INTRODUCTION

In reinforcement learning problems, agents take sequential actions with the goal of maximizing a time-delayed reward. In this chapter, the design of reward shaping for a continuing task in a multiagent domain is investigated. We use an interesting example, keepaway soccer (Kuhlmann, 2003; Stone, 2002; Stone, 2006), in which a team tries to maintain ball possession by avoiding the opponent’s interceptions. The keepaway soccer problem, originally suggested by Stone (2005), provides a basis for discussing various issues of multiagent systems and reinforcement learning problems(Stone, 2006). The difficulties of this problem are twofold, i.e., the state space is continuous and the sense-act cycle is triggered by an event, such as a keeper (learner) getting the ball. Since the learner selects a macro-action which

DOI: 10.4018/978-1-60566-898-7.ch013
requires a different time period, it is appropriate to model this problem as a semi-Markov decision process.

To our knowledge, designing the reward function has been left out of reinforcement learning research, even though the reward function introduced by Stone (2005) is commonly used. However, designing the reward function is an important problem (Ng, 2000). As an example, the following are difficulties of a designing reward measure for keepaway. First, it is a continuing task that has no explicit goal to achieve. Second, it is a multiagent cooperative task, in which there exists a reward assignment problem to elicit desirable teamwork. Because of these two features of keepaway, it is hard to define the reward signal of each keeper to increase the time of ball possession by a team. It should be noted that the reward for increasing each keeper does not always lead to increased possession time by a team.

In the case of a continuing task, we can examine a single-agent continuing task such as the pole balancing task, in which one episode consists of a period from the starting state to the failure state. If the task becomes a failure, a penalty is given, and this process can be used to evaluate teamwork and individual skills. In contrast, in the case of a multiagent task, which includes both a teammate and at least one opponent, it is hard to tell who contributes to the task. In a multiagent task such as keepaway, it is not always suitable to assign positive rewards to agents according to the amount of time cycles of each agent. Appropriately assigning an individual reward for each agent will have a greater effect on cooperation than sharing a common reward within the team. But, if the individual reward is not appropriate, the resulting performance will be worse than that after sharing a common reward. Therefore, assigning an individual reward to each agent can be a double-edged sword. Consequently, our focus is on assigning a reward measure that does not have a harmful effect on multiagent learning.

The rest of this chapter is organized as follows. In the next section, we describe the keepaway soccer domain, and discuss its features from the viewpoint of reinforcement learning. In Section 3, we introduce the reinforcement learning algorithm we applied and our reward design for keepaway. Section 4 shows our experimental results, including the acquired behavior of the agents. In Section 5, we discuss the applicability of our reward design on reinforcement learning tasks. We state our conclusion and future work in Section 6.

PROBLEM DOMAIN

Keepaway Soccer

Keepaway (Stone, 2002) is known as a subtask of RoboCup soccer, and it provides a great basis for discussion on important issues of multiagent systems. Keepaway consists of keepers who try to keep possession of the ball, and takers who attempt to take possession of the ball within a limited region. The episode terminates whenever takers take possession or the ball runs out of the region, and then players are reset for a new episode. When takers keep the ball for more than four cycles of simulation time, they are judged to have gained ball possession successfully.

Figure 1 shows the case of three keepers and two takers (3 vs. 2) playing in a region of size 20×20[m]. Here, keeper $K_1$ currently has the ball, $K_2$ is the closest to $K_1$, and $K_3$ is the next closest, and so on, up to $K_n$ when $n$ keepers exist in the region. In a similar way, $T_1$ is the closest taker to $K$, $T_2$ is the next closest one, and so on, up to $T_m$, when $m$ takers exist in the region.

Macro-Actions

In the RoboCup soccer simulation, each player executes a primitive action, such as a turn (angle), dash (power) or kick (power, angle) every 100[ms].
Related Content

Tropos: A Requirements-Driven Methodology for Agent-Oriented Software
[www.igi-global.com/chapter/tropos-requirements-driven-methodology-agent/5054?camid=4v1a](www.igi-global.com/chapter/tropos-requirements-driven-methodology-agent/5054?camid=4v1a)

Assessing the Impact of Temperature Change on the Effectiveness of Insecticide-Treated Nets
[www.igi-global.com/chapter/assessing-impact-temperature-change-effectiveness/66034?camid=4v1a](www.igi-global.com/chapter/assessing-impact-temperature-change-effectiveness/66034?camid=4v1a)

A Two-Layer Approach to Developing Self-Adaptive Multi-Agent Systems in Open Environment

The Human Mirror Neuron System
[www.igi-global.com/chapter/human-mirror-neuron-system/19632?camid=4v1a](www.igi-global.com/chapter/human-mirror-neuron-system/19632?camid=4v1a)