Incredibles 2006 Report

Atefeh Mirsafian[1], Firoozeh Sepehr, Samira Rezaee, Reyhaneh Movahhed, Mahdi Heidari, Ehsan Omidi, Sepehr Shojaee and Amir Hossein Shakoory

[1] mirsafian@ce.sharif.edu

Abstract. Incredibles is a group of high school students working on AI concepts using RoboCup simulation test-beds from 2004. The groups is currently working on soccer 2D and rescue simulation environments. The soccer team is based on Mersad team and after developing some new skills and plans, the whole team is trying to use reinforcement learning methods and neural networks to optimize its base.

1 Introduction

Incredibles is a group of high school students working on AI concepts. The team is based on Mersad soccer simulation team. Implementing some new skills and using AI algorithms to optimize some of the existing skills and to decide which plans to run, are two major tasks the team is working on. In section 2, one of the skills, called safe pass is described. In section 3, we will describe the learning method used to decide about plans.

2 Safe Pass

One of the most important skills of a football player is the pass skill. Normally, there are some mechanisms that can estimate how safe a pass may be. These mechanism usually use the current knowledge of agent about the world, and try to guess what happens if a special command is performed at next cycle. The output of these mechanisms is usually the probability of success of the pass, and maybe calculated based on different parameters, for example, the time difference of intercepting the ball by teammates and opponents.

The key idea in safe pass is to let the target of the pass act abnormal for some cycles, for example, move exactly in different direction, move to its strategic position, or stops searching for ball.

The maximum time for each abnormal action, and the probability that the agent acts so for exactly that time, depends on the team behavior an strategy. Soccer coach is really a powerful and useful tool to estimate this.

Suppose there are two kinds of abnormal actions, named A and B. Action A may take long up to 5 cycles and B may take long up to 3 cycles. Suppose the probability of happening A or B for sometimes is according to the following:

$$P_A[5] = \{ 0.1, 0.15, 0.35, 0.3, 0.1 \} \quad (1)$$

$$P_B[3] = \{ 0.15, 0.4, 0.45 \} \quad (2)$$

Suppose $S_A[i]$ and $S_B[i]$ are the success probability of a certain pass when agent is in state A or B respectively. Now, the success probability of the pass - when some abnormal actions is taken by teammates - can be estimated as follows:

$$S(A) = \sum_5 P_A[i]*S_A[i] \quad (3)$$

$$S(B) = \sum_3 P_B[i]*S_B[i] \quad (4)$$

And at last, the final success probability of the pass can be calculated based on the probability of the abnormal behavior of the agent easily:

$$P = P(A)*S(A) + P(B)*S(B) + P(N)*S(N) \quad (5)$$

Where P(A) and P(B) refer to the probability that an agent's behavior is either A or B respectively, and N is the normal behavior.

3 Using Q -Learning to Optimize Selecting Plan

Selecting among different plans in Mersad base code is hard-coded and based on different environmental conditions. We are going to use Q-Learning to optimize this. Q-Learning is one of the reinforcement learning methods. There is a decider component in the system whose action is to select one of the available plans. This decider updates itself based on the state of the world and the feedback that the simulation environment sends back to the algorithm.

As there are many states in this environment, each state in the Q-Learning algorithm is an index of some states in the real world. The award functions is the goal difference average during running this plan. For example, if running plan A leads to two goals per 1000 cycles, its award would be 0.002. The main step in Q-Learning algorithm - which is meant to be a recursive algorithm – is as follows:

$$Q(s, p) := Q(s, p) + a * (r + b*\max_q *Q(t, q-Q(s, p)) \quad (6)$$

Where, s is the current state of the world, p is one valid action for agent, Q(s, p) is the value of action p in state s, r is the award for taking action p in state s, t is the state of the world after performing a on s, and q is any valid action at t. The more the b is, the system will mostly pay attention to awards that will be collected in the future. a is also called as the learning coefficient.

Q-Learning will finally select the action a among all actions whose Q(s, a) is the most, which means selecting the action that can lead to more award.

4 Summary

Soccer simulation is a suitable test-bed for multi-agent AI algorithms. Mersad, has provided a multi-layer base for this environment that makes testing different approaches really easy. In this paper, we introduced Incredibles team and summarized some of these approaches.

References

1. The RoboCup Federation: The official RoboCup website at http://www.robocup.org
2. E. Foroughi, F. Heintz, et al: RoboCup Soccer Server User Manual for Soccer Server Version 7.06 and later, 2001. At http://sourceforge.net/projects/sserver
3. Meisam Vosoughpoor, Ali Boorghani, et all: Mersad 2005 Team Description