Fifty-Storms: Team Description 2010

Harukazu Igarashi[1], Jun Masaki[2], Norihiro Nagae, Akifumi Yoshimoto

Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan
{[1]arashi50, [2]*l*06103}@shibaura-it.ac.jp

Abstract. Fifty-Storms is a team that is participating in the RoboCup Soccer Simulation 2D League 2010. This team is based on open-version codes of the HELIOS team, agent2d, to which some modifications were added to enhance the offensive abilities of the forwards and defenders by hand coding. In addition, it exploits the results of reinforcement learning, where a policy gradient method derives appropriate policies for pass selection.

1 Introduction

Fifty-Storms was established in 2008 for RoboCup Soccer Simulation 2D League by members of a student project at the Shibaura Institute of Technology. The members of the student project are changed every year. From 2005, teams comprised of past members had participated in domestic competitions, RoboCup Japan Open, but they failed to succeed or did not participate in any international RoboCup competition. However, the new members changed their underlying base team from *UvA Trilearn 2003*[1] to *HELIOS*[2]. After adding some new tactical skills and making modifications to optimize HELIOS's abilities, they named their team Fifty-Storms for their adviser, Prof. H. Igarashi, whose name means "fifty storms" in Japanese. Fifty-Storms participated in RoboCup Japan Open 2008 and narrowly lost the championship match to HELIOS.

Next year, Fifty-Storms 2009 exploited the research results of H. Fukuoka, N. Sano, and H. Igarashi who applied reinforcement learning to pass selection problems of midfielders (MFs) and positioning problems for forwards (FWs) to receive a through pass. In their work, they proposed an objective function, which is a linear combination of heuristic functions evaluating an agent's action, and determined the values of weight coefficients in the objective function by a kind of reinforcement learning called the policy gradient method.

This year, Fifty-Storms 2010 applies the reinforcement learning to adjusting weight parameters in the objective function that HELIOS uses for pass selection. In this team description paper, we briefly describe the modifications added to HELIOS and the learning method.

2 Modifications Added to Base Team

Fifty-Storms uses HELIOS whose source codes were released by Hidehisa Akiyama and put on a URL site in [2] as *agent2d*. He also released a large number of useful functions as a software library called *librcsc* that are necessary for making soccer agent programs. Moreover, the source codes of HELIOS and detailed descriptions of the agent's skills and the team's tactics and strategies were combined and published in 2006 (in Japanese). After this book's publication, many Japanese teams changed their base team from UvA to HELIOS as Fifty-Storms did.

In Fifty-Storms 2009, enhancing offensive power is the main design principle to modify HELIOS. For that purpose, the following two ideas were designed and implemented. The first maximized the dribbling skill of agents in HELIOS so that a FW or a MF runs through the defense line of the opponent's defense players (DFs) and shoots as near the opponent's goal as possible. The original HELIOS had offending strategy characteristics in which a FW in a corner area in the opponent's field passes to teammates in front of the opponent's goal. That is called a side attack or an open offense. Fifty-Storms 2009 added another useful option in HELIOS's offense strategy.

The second idea takes an aggressive defense strategy by making DFs go to opponents with the ball so that DFs can take the ball from them. In addition, FWs do not return to their own field and instead stay near the opponent DFs, even if the opponent goalie catches the ball and kicks it. FWs add pressure on the opponent's DFs and try to take the ball from them.

3 Learning of Soccer Agents

3.1 Policy Gradient Approach to Soccer Agents

The RoboCup Simulation 2D League is recognized as a test bed to learn coordination in multi-agent systems because there is no need to control real robots and one can focus on learning coordinative behaviors among players. As an example of multi-agent learning in a soccer match, Igarashi et al. proposed and applied a policy gradient approach to realize coordination between a kicker and a receiver in direct free kicks [3]. They dealt with a learning problem between a kicker and a receiver when a direct free kick is awarded just outside the opponent's penalty area. They proposed a function that expressed heuristics to evaluate a candidate target point for effectively sending/receiving a pass and scoring. However, they only dealt with the attacking problems of 2v2 (two attackers and two defenders), and their base team used in [3] was UvA Trilearn 2003. They applied the policy gradient approach to pass selection problems of MFs and positioning problems for FWs to receive a through pass and implemented the learning results into Fifty-Storms 2009.

3.2 Characteristics of Policy Gradient Method

A policy gradient method is a kind of reinforcement learning scheme that originated from Williams's REINFORCE algorithm [4]. The method locally increases and maximizes the expected reward per episode by calculating the derivatives of the expected reward function of the parameters included in a stochastic policy function. This method, which has a firm mathematical basis, is easily applied to many learning problems. One can use it for learning problems even in non-Markov Decision Processes [5][6].

The policy gradient method used in refs. [3] and [6] has the following technical characteristics. For the autonomous action decisions and the learning of each agent, the policy function for the entire multi-agent system was approximated by the product of each agent's policy function [7] defined by

$$\pi_\lambda\left(a_\lambda; s_\lambda, \left\{\omega_j^\lambda\right\}\right) \equiv \frac{e^{-E_\lambda\left(a_\lambda; s_\lambda, \left\{\omega_j^\lambda\right\}\right)/T}}{\sum_a e^{-E_\lambda\left(a; s_\lambda, \left\{\omega_j^\lambda\right\}\right)/T}}, \tag{1}$$

where a_λ is the action of agent λ and s_λ is the state perceived by agent λ. Function E_λ in (1) is an energy function of Boltzmann distribution function π_λ and an objective function that evaluates an action of agent λ. At the end of each learning episode σ, common reward $r(\sigma)$ is given to all agents. The derivative of expectation of reward $E[r]$ for parameter ω_j^λ can be calculated to derive the following learning rule on ω_j^λ:

$$\Delta\omega_j^\lambda = \varepsilon \cdot r(\sigma) \sum_{t=0}^{L(\sigma)-1} e_{\omega_j}^\lambda(t) \Big/ T, \tag{2}$$

where $L(s)$ is the length of episode σ and ε (>0) is a learning coefficient. e_ω are called characteristic eligibilities[4] and shown in (3) when π_λ is given by (1).

$$e_{\omega_j}^\lambda(t) \equiv \frac{\partial}{\partial\omega_j} \ln\pi_\lambda = -\frac{1}{T}\left(\frac{\partial E_\lambda}{\partial\omega_j} - \left\langle\frac{\partial E_\lambda}{\partial\omega_j}\right\rangle_{\pi_\lambda}\right), \tag{3}$$

where $\langle X\rangle_\pi$ means the expectation of $X(a)$ with respect to stochastic variable a distributed with distribution function π.

4 Pass Selection Problem

4.1 Policy Gradient Approach to Soccer Agents in Fifty-Storms 2009

By applying the policy gradient method summarized in Section 3, Fifty-Storms 2009 exploits the learning results obtained to pass selection problems of MFs and positioning problems for FWs to receive a through pass. We used E_λ in (4) that is a

linear combinaiton of heuristics functions $U_j(a_\lambda;s_\lambda)$ that seem to be useful for selecting a pass receiver or an appropriate receiving point.

$$E_\lambda\left(a_\lambda;s_\lambda,\left\{\omega_j^\lambda\right\}\right)=-\sum_j \omega_j^\lambda \cdot U_j\left(a_\lambda;s_\lambda\right). \quad (4)$$

In Fifty-Storms 2009, five heuristics from U_1 to U_5 for evaluating the current position of a teammate as a desirable pass receiver. U_1, U_2, and U_3 are heuristics for passing the ball safely. U_4 is a heuristics for making an aggressive pass. U_5 is used for treating reliable information as more important knowledge than unreliable information. All U_i's are normalized between 0 and 10.

4.2 Policy Gradient Approach to Soccer Agents in Fifty-Storms 2010

In Fifty-Storms 2009, we used our original objective function described in 4.1 for a passer to select a teammate to pass the ball or for a receiver to select a point for receiving the pass. However, agent2d has the following objective function for a passer to select a target point on the field for his pass.

$$E_\lambda(a_\lambda;s_\lambda,\{\omega_j^\lambda\})=-\prod_{j=1}^{M} U_j(a_\lambda;s_\lambda,\omega_j^\lambda), \quad (5)$$

where

$$U_j(a_\lambda;s_\lambda,\omega_j^\lambda)=\omega_j^{\alpha_j(a_\lambda;s_\lambda)}. \quad (6)$$

$\alpha_j(a_\lambda;s_\lambda)$ evaluate agent's action a_λ at state s_λ. $\omega_j(\in[0,1])$ are the weight parameters of the j-th heuristics. If we use (5) and (6) instead of (4), the policy gradient method gives different learning rules from those used in Fifty-Storms 2009. In Fifty-Storms 2010, the following rules are used,

$$\Delta\omega_j^\lambda=\varepsilon\cdot r\cdot\frac{1}{T}\sum_{t=0}^{L-1}\left[E_\lambda(a_\lambda;s_\lambda)\alpha_j\left(a_\lambda;s_\lambda\right)/\omega_j^\lambda-\left\langle E_\lambda(a_\lambda;s_\lambda)\alpha_j\left(a_\lambda;s_\lambda\right)/\omega_j^\lambda\right\rangle_{\pi_\lambda}\right]\cdot \quad (7)$$

5 Summary

In this team description paper, we outlined the characteristics of the Fifty-Storms team that is participating in the RoboCup 2D Soccer Simulation League 2010. This team is based on an open version of the HELIOS team, agent2d (ver.2.1.0), and to which some modifications to HELIOS were added to enhance the offensive abilities of forwards and defenders by hand coding. In addition, it exploits the results of reinforcement learning, where a policy gradient method derives appropriate policies for pass selection.

References

1. UvA Trilearn 2003, http://staff.science.uva.nl/~jellekok/robocup/2003/
2. HELIOS's URL site (in Japanese), http://rctools.sourceforge.jp/pukiwiki/
3. Igarashi, H., Nakamura, K., and Ishihara, S.: Learning of Soccer Player Agents Using a Policy Gradient Method: Coordination between Kicker and Receiver during Free Kicks. In: 2008 International Joint Conference on Neural Networks (IJCNN 2008), Paper No. NN0040, pp. 46-52 (2008).
4. Williams, R. J.: Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. Machine Learning, vol. 8, pp. 229-256 (1992).
5. Igarashi, H., Ishihara, S., and Kimura, M.: Reinforcement Learning in Non-Markov Decision Processes-Statistical Properties of Characteristic Eligibility-. IEICE Transactions on Information and Systems, vol. J90-D, no. 9, pp. 2271-2280 (2007, in Japanese). This paper is translated into English and included in The Research Reports of Shibaura Institute of Technology, Natural Sciences and Engineering, vol. 52, no. 2, pp. 1-7 (2008). ISSN 0386-3115
6. Ishihara, S. and Igarashi, H.: Applying the Policy Gradient Method to Behavior Learning in Multiagent Systems: The Pursuit Problem. Systems and Computers in Japan, vol. 37, no. 10, pp. 101-109 (2006).
7. Peshkin, L., Kim, K. E., Meuleau, N., and Kaelbling, L. P.: Learning to Cooperate via Policy Search. In Proc. of 16th Conference on Uncertainty in Artificial Intelligence (UAI2000), pp. 489-496 (2000).