DreamWing2D Simulation 2D Team Description

Paper for RoboCup 2012

Huilong Zhang, Fajun Zhao, Yanan Fan, Changning Huang,
Yu Gan, Binbin Ruan

Department of Computer Science and Technology,
Anhui University, China
zhaofajun216@126.com

Abstract. This paper briefly describes the background DreamWing2D 2012 and recent research results of Dreamwing2D 2012. First, we apply the Q-learning to the base of the agent's intelligence assessment system, so that the player agent's action can gradually approach the optimal strategy of action. Furthermore, through enhancing the cooperation between the side back players and the side forward players, they can intercept the ball easily and prevent the opponents from dribbling and passing. Besides, we have modified other aspects, such as formation, marking, etc.

1. Introduction

Simulation 2D soccer team of Anhui University was founded in 2006, and made the top 24 results in the RoboCup ChinaOpen 2006. In 2007, we were the top 8 teams in the Robocup ChinaOpen 2007. In 2009, we reached 2nd place in the RoboCup 2009 of Anhui Province. Then, we took part in the RoboCup ChinaOpen 2009 and got the 12th place of soccer simulation 2D. In the RoboCup ChinaOpen 2010, we got the 7th place. And in the RoboCup ChinaOpen 2011, DreamWing2D achieved a breakthrough in history and won the 3rd place.

Dreamwing2D 2012 inherits from Dreamwing2D 2011 which is based on released code of Agent2D-3.1.0, but she has a significant improvement. Although our team is not powerful enough, we are becoming stronger and stronger. We hope to gradually improve ourselves by participating in more games.

2. Q-learning algorithm

2.1 Frame of agent2D

DreamWing2012 is based on the concept of chain of actions which is put forward in Agent2D[1], through which the agent can make the next a few cycles of action in series instead of only deciding what to do for the next one cycle. Through the Q-learning algorithm, the agent can choose each action of every chain of actions. That is to say each cycle agent takes max Q value as the best action.

The framework of agent2D is shown in Figure 1.

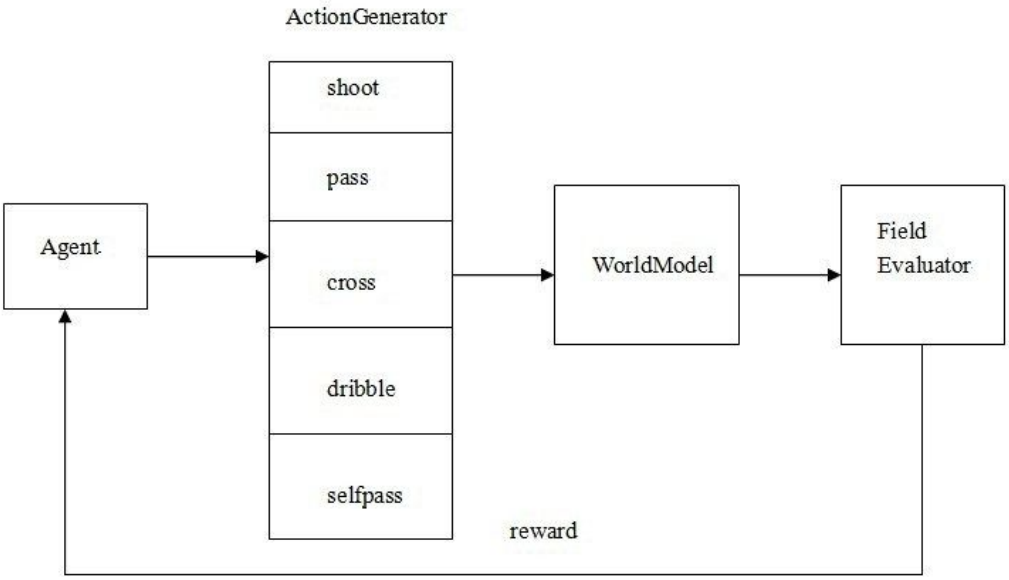

Fig.1. Strategy-making framework of agent2D

2.2 Q-learning

As mentioned before, the choice of actions of Dreamwing2D 2012 is achieved by Q-learning algorithm. In Fig.1, Dreamwing2D 2012 applies the Q-learning algorithm to the Field Evaluator, so that agent has a better performance in the choice of actions.

Q-learning algorithm mainly relate to three issues: the action set, the state set and the reward. [5,6,7,8]The description of the action set is 5 actions in ActionGenerator : shoot, pass, cross, dribble, selfpass. By the discrete of many states, Dreamwing2D 2012 takes a number of representative and significant real-time variables of pitch as state set we need. The calculation of the reward is the core of Q-learning algorithm.

The details can be described as follows:

(1) Our goals, reward=1

(2) Ball reaches the shooting points, rewards=0.9

(3) Ball out of pitch, reward=0

(4) None of the above, reward=basic reward of region (BRR) + reward within the region (RWR)

We integrate the various factors of front of pitch especially the penalty area and calculate the shooting points, through which agent can determine to shot or not. In (4), we divide the front of pitch into 30*10 small regions to get agent's position. In order to get reward, we divide the front of pitch into 4 large regions (shown as Fig.2.). The BRR of the region where the ball is is the current BRR.

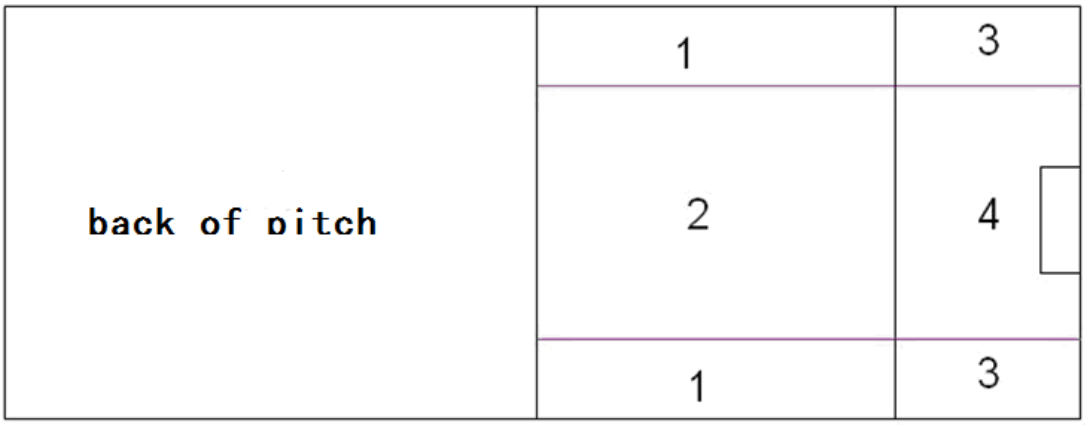

Fig.2. the divided of the front of pitch

In Fig.2, the same number represents that they have the same values. The 4th region has the max BRR because there is benefit to goal. The 3rd regions are followed. And then are 1st and 2nd regions. If the offensive team focused on the middle breakthrough, the value of the 2nd region is greater than 1st region. Or the value of the 1st region is greater than 2nd region. Dreamwing2D 2012 focuses on the side breakthrough.

In addition to BRR, we also need RWR to get accurate values. RWR is mainly defined by the information of the ball's position, nearest offensive player and defensive player. In (4), we must make sure that reward < 0.9 and BRR should be at least 10 times greater than RWR. The method of Dreamwing2D 2012 get RWR can be described as follows:

If d_1>5.0：RWR=(X_A+（d_1-5.0）*2.0*X_A-d_3)/100;

If 3.0<d_1≤5.0：RWR=(X_A+（d_1-d_2）*2)/100;

If d_1≤3.0：RWR=(X_A-d_3)/100-BRR;

In these Formulas, X_A is the ball's X coordinate, and d_1 is the distance between the defensive player and the ball. d_2 is the distance between the offensive player and the ball, and d_3 is the distance between the ball and the goal. Agent can control dribbling speed by controlling the value of X_A and decide whether shooting by controlling the value of d_3. By controlling the value of d_1 the agent can decide whether get rid of the nearest defender. Through RWR minus the current BRR, agent's reward decreases rapidly, and then the agent will pass or shoot to get out of the state quickly.

The formula of updating the Q value is $Q(s_t, a_t)=(1-\alpha)\ Q(s_t, a_t)+ \alpha(r(s_t, a_t)+\gamma \max Q(s_{t+1}, a))$[6,8], and α=0.15，γ=0.85. When the current situation is the final states, the agent will update Q value according to that formula. The final states include that we goals, the ball is out of pitch and the ball is intercepted by defensive player.

3. The Improvement of Formation and Skills

Thanks to Dreamwing2D 2012 uses Agent2D as the base code, so she also uses Agent2D's Formation [1,4]. However, through modifying of the basic formation, we have our own characteristics of the formation. For instance, the side back guard can block the opponent players and they cannot easily dribble through flank by making the position of guard closer to region of flank. Besides, the formation of the formation files is only a basic moving point. Dreamwing2D

2012 still modifies the moving points based on basic points such as marking. In order to implement marking in different play mode, the agent should move to the particular position which is based on the opponent players and don't change our defensive formation. In this way, when the opponents pass the ball, our players can intercept the ball easily.

Besides, we have also enhanced the back guard's ability of force snatch by enhancing two players' coordinating.[6] In Fig.3, before modifying, when the offensive player dribbles breakthrough from side, the defensive player doesn't intercept until it reaches the penalty area. But at this time they have threatened our goal, and they can goal easily. In Deamwing2D 2012, we implement force snatch through two players by modifying the players' moving points and improving the flexibility of the players. In Fig.3, the defensive player (num 4) seals the offensive player's (num1) route, and then the side forward player (num 3) force snatches from the flank. At the same time, the position of num 3 is between the position of num 1 and num 2. It prevents num1 from passing the ball to num 2. The statistical data show that in this way, the success rate of intercepting is very high.

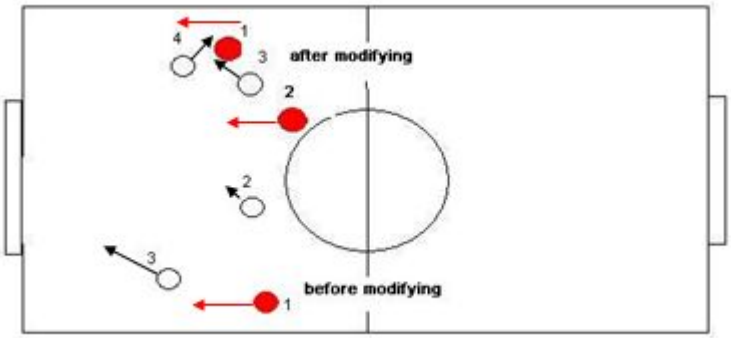

Fig.3. the schematic diagram of the effect of force snatch

4. Conclusion and Future Work

Through a large amount of training, the Q values gradually converge to stable vales in different states. After training, the agent is put into the games, according to statistical data of the number of goals, we get great results. Especially when we put the training scenarios into edges of penalty area, the agent can dribble through goalie and defender flexibly in small-scale after training, and greatly increase the number of shooting.

Table 1. The result of competing with some teams

	UVA		Helios_Base	
	Ave Goals Scored	Win	Ave Goals Scored	Win
Helios_Base	1.02	43.7%	2.14	44.3%
Dreamwing2D 2011	1.95	51.6%	3.02	56.7%
Dreamwing2D 2012	4.43	82.1%	5.25	84.3%

In table 1, we list goals and the percentage of winning which are Helios_Base, Dreamwing2D 2011 and Dreamwing2D 2012 competing with excellent code based on UVA and Helios_Base after 100 games. In table 1, we can see that Dreamwing2D 2012 has a great improvement so that

goals and winning have a leap.

Although Dreamwing2D 2012 has a great improvement related to the earlier version, she is not perfect on the whole. We will focus on the researching of the receiver when passing and completing the communication system in future work. We believe Dreamwing2D 2012 will have better performance in future games!

References:

[1] Hidehisa Akiyama and Hiroki Shimora. HELIOS2010 Team Description

[2] China University of Science and Technology, Design and Implementation of simulation robot soccer

[3] http://ai.ustc.edu.cn/2d/

[4] http://sourceforge.jp/projects/rctools/

[5] Yong Ma, Longshu Li, Xuejun Li. Q-learning based on Intelligent Agent research and application of defensive strategy. Computer calculation and development. 2008

[6] Kostiadis Kostas, Huosheng Hu. Reinforcement Learning and Co-operation in a Simulated Multi-agent System [A]. In: Proceedings of the 1999 IEEE / RSJ International Conference on Intelligent Robots and Systems 1999 [C] ,17-21 Oct. 1999, 2: 990-995.

[7] School of Information Science and Engineering. Enhance the Q study in non-Markov system to determine the application of optimization problem

[8] Jiawang Zhang, Guangsheng Han, Wei Zhang. Q-learning algorithm of the ball in the RoboCup. System simulation. 2005