HfutEngine2011 Simulation 2D Team Description Paper

Rongya Chen, Yiming Li, Hao Wang, Baofu Fang

Artificial Intelligence And Data Mining Lab

Abstract. This paper describes the background,the framework and the design feature of the HfutEngine2011. We put forward a new approach to do research on Multi Agent System. The method is based on mining teammate behavior.In this scene, an autonomous coach agent is able to get the current information of all teammates without noise, which can be modeled to compose patterns of teammates. At first coach agent gathers data from noisy environment to identify pattern of player agent.Then compute the probability of pattern compared with current situation by statistical calculations. Then the player agent analyzes the communication messages from teammates and the see message from server. The player agent decides the best message to choose and enforces the former behavior. Finally the player agent makes best decision according to the result of mining teammate behavior.

1 Introduction

Team HfutEngine was founded in 2002 and took part in the RoboCup ChinaOpen2002. In the following years,HfutEngine develops fast and joins many matches. From 2005, we use UVA_BASE_2003 as our base code, we've added our own AI methods to it and updated the code along with the server's upgrade. In RoboCup ChinaOpen 2007 we took the 2nd place of soccer simulation 2D. We took the 7th place of soccer simulation 2D in World RoboCup 2008 and 3rd place of soccer simulation 2D in RoboCup ChinaOpen 2010. We hope to obtain a good grade in World RoboCup 2011. We want to probe into Multi-Agent System and Robocup with anyone interested in them.

2 Framework of HfutEngine2D

In our exploitation we found that any Multi-Agent cooperation was based on how the single agent adapt to the Multi-Agent System. If every element in the system can accommodate with the system, the system is steady. There is not need to make unitive command for every agent. Our strategy is based on the value judgment, that is every agent has its own evaluator to calculate correspond value. Then the action which has max value is being executed by executant.

This framework first founded in 2005. The former methods make use of evaluator to design to do something like shoot, dribble, etc and then to perform

the relevant action. In this way, the decision of the evaluator mainly depend on experience knowledge. Now we depend on the product of income and the probability of success which are estimated by environment beforehand to decide the action. The idea is not easy to perform. The player decide to perform an action by the balance of the value.

This structure is shown in Figure 1.

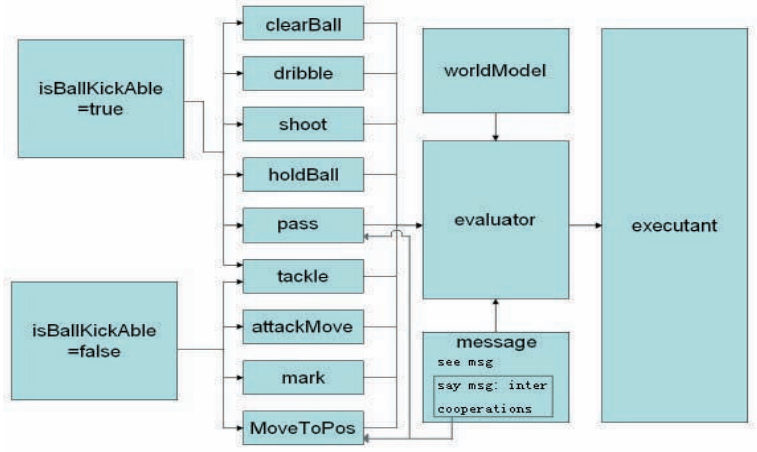

Fig. 1. The structure of HfutEngine2D.

3 The High Level of HfutEngine2011

The HfutEngine2011's high decision includes two parts, evaluator and executant. Evaluator predict the benefits token by all the actions, then we can Make a decision which action should be executed. Executant take responsible for how to execute the action.

Firstly, online coach agent gathers information by visor from environment,then constructs model of teammates, using χ^2 test to identify other teammate patterns. Meanwhile, coach send new model to server during the game. Teammates get new model of others from server and use Q-learning to decide current measure. Finally, coach feeds new information back to surroundings. Figure 2 shows the diagram of mining teammate behavior.

3.1 Evaluator

The main application of evaluator is to make best decision. Because the environment is dynamic and stochastic, we can assume every action as random distribution. So we assume the decision of agent in the whole game as Gaussian distribution. Each agent makes decision independently. We all know that

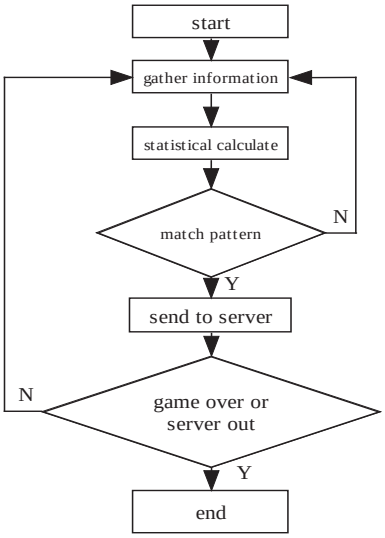

Fig. 2. Diagram of Mining Teammate Behavior.

sum of Gaussian distribution is χ^2 distributing. In this paper we make use of χ^2 distributing to get the agent's pattern. For example, the ball is accelerated by agent A, and later received by agent B. This pattern can be considered as passing ball.

The step of evaluator can be described as follows:
Step1: We gather all the information of teammates from surroundings and check the current intention of agent.
Step2: We use following function to compute whether the current pattern is passing ball.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

Attention: O_i means effect passing times, E_i means expect passing times
Step3: Use Q learning to train best Q table based on pattern computed by step2.
Step4: Match passing ball pattern. We HfutEngine2D adopt formation523, and we take player2 pass ball for example, and the result shows in Table 1:
Step 5: Analyze the matching result. We assume $\alpha \equiv 0.1$, which means the ratio of pass ball is ≥ 0.9. We get 6.251 from distributing table. The computing result is 3.596, we know that $3.596 \leq 6.251$, so we can confirm the feasibility of pass ball pattern is ≥ 0.9. Now we decide that the intension of agent is to pass ball. Figure 3 show the relationship between pass ball matching probability and training time.

Table 1. passing ball result

pass times	num 6	num 7	num 8	num 10	sum of pass times
effect passing times	93	454	172	181	1000
expect passing times	80	470	280	170	1000
Result	2.11	0.545	0.23	0.711	3.596

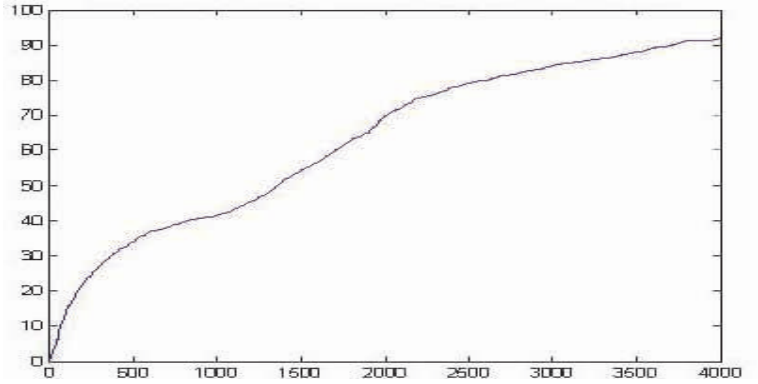

Fig. 3. Relationship Between Pass Ball Matching Probability and Training Time

3.2 Executant

The executant includes how the player execute the actions. Some of them are gained by training, others are deduced by the mathematic and physical expressions combined with the rules of the server. Take shoot for example.

The action shoot is obtained by training. The training scene can be seen at left part of Figure 4. The arithmetic is as follows:

(1)Fix the position of the ball and the shooting-player. Set the position of the goalie randomly and record the position.

(2)If the shooting-player take the ball, then shoot to the fixed position in the goal.

(3)If the ball is kicked in the goal, then record the shoot is success ,whereas if the ball is hold up by the goalie then record the shoot is failure.

Repeat the (1) to (3),we can get a set of value of posGoalie, base on the value of the 'posBall' and 'posShootingPoint', then we can compute the α and d. We use network of Radial Basis Function to train these values. Make the Gauss Nucleus Function and network framework as two inputs and one output. We have 3 parameters to learn including the center of Radial Basis Function, the weight from center-cell to output-cell, variance. The training result show that the ratio of success is up to 91.28% compared to 58.18% . Shoot-Advisor need to submit Method to the higher framework. The method need 3 parameters.

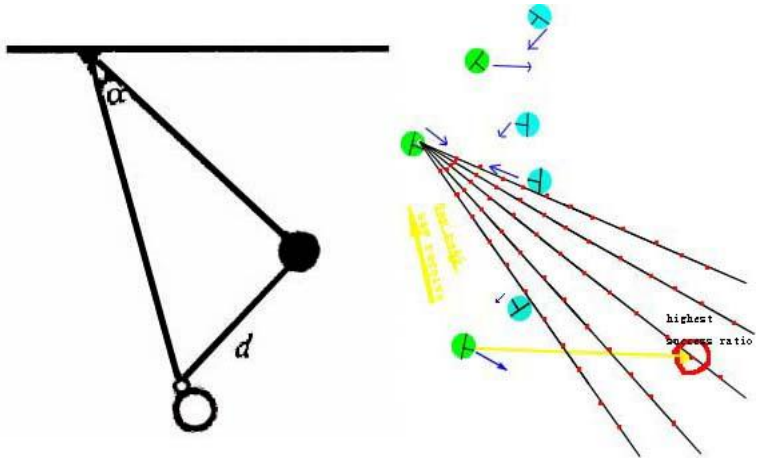

Fig. 4. New Pass Ball Model

3.3 New pass-ball Model

We have update our pass-ball model.Now the model calculates success ratio of more points(effective points,these points are on several lines) and uses more effective method, it considers only the destination in former versions. First,our model calculates the largest space to pass. Then our model divide the space to several angles with lines(the amount of angles dependence on the speed of the machine, we trained to get the amount). Finally it gets some points on the line for calculation.We plan to use this model by offline training later,because the method requires too many system resource.

Right part of Figure 4 shows this new pass-ball model.

3.4 New Communication System

The exact location of objects on the pitch, the decision has always been the key to the world's model.The combination of the visual information and auditory information has been an important method to improve the level of our team.

We improved our communication system. When the players can not see some objects he will obtain information through auditory information. The auditory information is useful when a player is serious lack of visual information. Through the auditory information the player can get more information of invisible area, and through it players can manage better cooperations. This improved communication system improves the ratio of pass success.

And we take the advantage of the coach.When a player fouls,gets a card or is of low stamina,the coach's substitutions takes great effect.

3.5 Dynamic Positioning and Role Exchange

Dynamic positioning and role exchange(DPRE) is based on previous work that suggested the use of flexible agent roles with protocols for switching among them. Players may exchange not only their positions(place in the formations) but also their player types in the current situation. Positioning exchanges are performed only if the utility of that exchange is positive for the team. Utilities are calculated using the player's positions to their strategic positions, the importance of each positioning and the adequacy of each positioning in the formation on that situation

3.6 Reinforcement Learning

All involve interaction between an active decision-making agent and its environment, within which the agent seeks to achieve a goal despite uncertainty about its environment. The agent's actions are permitted to affect the future state of the environment(e.g., the next ball position, the formations, the next agent position), thereby affecting the options and opportunities available to the agent at later times.Correct choice requires taking into account indirect, delayed consequences of actions, and thus may require foresight or planning..
At the same time, the effects of actions cannot be fully predicted, thus the agent must monitor its environment frequently and react appropriately. The agent can use its experience to improve its performance over time. The knowledge the agent brings to the task at the start–either from previous experience trained offline or built into it in the game.
Whereas a reward function indicates what is good in an immediate sense, a value function specifies what is good in the long run.

$$\sum_{j \in \mathbf{N}} b_{ij}^{(\lambda+1)} \hat{y}_j = \sum_{j \in \mathbf{N}} b_{ij}^{(\lambda)} \hat{y}_j + (b_{ij} - \lambda_i)\hat{y}$$

It is values with which we are most concerned when making and evaluating decisions. Action choices are made based on value judgments. We seek actions that bring about states of highest value, not highest reward, because these actions obtain the greatest amount of reward for us over the long run.

4 Conclusion and Future Works

The practice prove that the design idea of value-judge is very successful. The achievement that we have got in the past year further improve that point and the ability of our team make a great leap. In Table 2, we can see the result of competing with some teams. It shows that HfutEngine2011 with this design have good match-ability.

We hope that we can improve more segments of our team. We will further optimize this framework and exert it's flexibility advantage. We will solve the problem of our team gradually.Also, we plan to make a new way to forecast

Table 2. the result of competing with some teams

Team	Ave Goals Scored	Ave Goals Conceded	win	draw	lose
HfutEngine2007	8.9	0.19	20	0	0
HfutEngine2008	5.9	0.13	20	0	0
HfutEngine2009	2.0	0.50	15	4	1
HfutEngine2010	1.8	0.40	17	2	1
Mersad2005	3.7	0.43	17	2	1

the situation in the playground based on experience which include the states of teammates, opponents, ball and so on. Based on this new method we can get the value more correctly.We also continue to make research on the Multi-Agent System and Machine Learning in order to enlarge the ratio of learning in our team.Meanwhile, the research will be focus on the fast-online learning not the off-line accumulated learning. The fast-online learning make the player learn to change on-time.In the coming time we will work hard to make a good result in the World RoboCup.

References

1. Xiaoping Chen, et al, Challenges in Research on Autonomous Robots, Communications of CCF, Vol. 3, No. 12, Dec 2007.
2. Kitano H, Tambe M,Stion P,et al. The Robocup synthetic agent challenge 97[A].Procceeding s of Fifteenth International Joint Conference on Artificial Intelligence (IJCA-97)[C]. Berlin:Springer,1997.24-29.
3. Peter Store, Tucker Balch, Gerhard Kraetzschmar(Eds.).RoboCup 2000: Robot Soccer World Cup IV. Mar, 2002.
4. Baofu Fang,Hao Wang ,ect.The Survey of HfutEngine2005 Robot Simulation Soccer Team Design, Journal of Hefei University of Technology. Vol.29(9). Sep, 2006.
5. Gao Jian-qing, Wang Hao, Yu Lei. A Fuzzy Reinforcement Learning Algorithm and Application in RoboCup Environment. Computer Engineering and Applications, 2006(6).
6. Yang LIU, Hao WANG, Baofu FANG, Hongliang Yao. Application of the method of supp vector regression in RoboCup. Journal of Hefei University of Technology Vol.30(10). Oct,2007.