Soft Computing (2022) 26:10939–10946
https://doi.org/10.1007/s00500-022-07210-9

FOCUS

Football training evaluation using machine learning and decision support system

Qiangqiang Xu[1] · Xin He[2]

Accepted: 6 January 2022 / Published online: 20 July 2022

Abstract
With the rapid development in information and communication technologies, the relationship among the disciplines in various fields has become closer, resulting in many new technologies based on interdisciplinary research. This article considers robot soccer training as a research topic by involving related fields, i.e., interdisciplinary research. Compared with the physical robot, its simulation has the advantages of less cost, shorter experimental period, and is beneficial to the real-life studies in this context. This article discusses the practical application and optimization techniques used for robot soccer matches and training evaluation. The main tasks are as follows. First, we analyze and build a simulation platform for robot soccer matches. The basic principles of this platform are learning, and designing a novel optimized SARSA algorithm, built using the state-of-the-art state–action–reward–state–action (SARSA) algorithm. The optimized SARSA is a reinforcement learning algorithm that uses Q-value for making critical decisions about various parameters involved in the football training. Second, this paper simulates and analyzes the proposed optimized SARSA algorithm in a single-entity environment of robot football training. We compare the two algorithms before and after the improvement in a multi-entity environment. Among the simulation results, it is found that the optimized SARSA has a more powerful performance. The simulation results prove that the robot football training evaluation based on machine learning (optimized SARSA) can better formulate the robot football training strategy after applying the reinforcement learning strategy.

Keywords Artificial intelligence · Machine learning · Football training · Decision support system · SARSA

1 Introduction

Machine learning is a multi-disciplinary field that involves many disciplines such as statistics, probability theory, convex analysis, approximation theory, and algorithm complexity theory (Wiley et al. 2003; Lu and Li 1999; Marjanovi et al. 2011). Machine learning theory mainly involves designing and analyzing some algorithms that allow computers to learn independently. In general, a machine learning algorithm is an algorithm that automatically analyzes and obtains values from data and uses the accepted rules to predict unknown data. Since many statistical theories are involved in learning algorithms: machine learning and statistical inference being particularly closely related, it is also called statistical learning theory (Zheng et al. 2003; Vidyasagar 2015). One well-known and popular application of machine learning is the robotic soccer, which uses an interdisciplinary research and is extremely interesting, admirable, and famous, so it has attracted more and more attention. For experts and scholars, the robotic football game is an excellent platform for scientific research and creation, especially in artificial intelligence (Dewey et al. 2015). Over the last few decades, artificial intelligence research has mainly focused on solving problems in a single-agent, static, and predictable environment (Zhao 2007). One common example is chess man–machine competition. With the development of artificial intelligence, people began to pay attention to the research of multi-agent issues. In the future, the main

Communicated by Tiancheng Yang.

✉ Xin He
10048@buaa.edu.cn

Qiangqiang Xu
XQQ@dhu.edu.cn

1 Department of Physical Education, Donghua University, Shanghai, China

2 Department of Physical Education, Beihang University, Beijing, China

problem of artificial intelligence will be the problem-solving of multi-agent dynamic and unpredictable environments. The multi-agent system of various agents is a microcosm of the future world (Dimeas and Hatziargyriou 2005). How to coordinate and cooperate between agents is the fundamental problem of multi-agent research, such as agent design, multi-agent architecture, automatic reasoning planning, and machine learning (Elliott 1992). In this context, a series of issues such as knowledge acquisition have been intensively reflected in the robot football.

The robot football can be regarded as the standard problem for the convergence of artificial intelligence and robotics in the future (Le et al. 2008). The robot football platform is a perfect experimental platform for researchers, through which many new algorithms and new methods can be experimented with and tested. Researchers can observe the efficiency of the algorithm and experimental results during the test to find the deficiencies of the algorithm and make improvements. One can also try to synthesize the advantages of existing algorithms to improve the execution effect of the agent. In addition to promoting the development of theoretical research in the field of artificial intelligence, the research of football robots can also promote the better application of artificial intelligence research results in real life. For example, the study of autonomous football robots can be applied to the direction of unmanned combat vehicles, while dynamic shooting is similar to missile interception, and the design of defensive strategies can be applied to anti-tank missiles (Xu 2004). In addition to military applications, the research directions of soccer robot games can also be applied to home entertainment, business management, and other aspects.

The main research content of this paper is the application of reinforcement learning on the MSRS football robot simulation platform. First, the basic principles of reinforcement learning and several classic algorithms are studied, and then, the design process and experimental results of the application of reinforcement learning in single-agent and multi-agent environments are described, respectively. The main objective is to discuss the practical application and optimization methods of machine learning in robot soccer matches and training evaluation, and conduct simulation experiments. The results of the experiments revealed the effectiveness of the proposed study.

The rest of this paper is organized as follow. In Sect. 2, the related work is presented. In Sect. 3, the proposed methodology using reinforced learning model for football training evaluation is presented. The experimental results and discussions are provided in Sect. 4. Finally, the paper is concluded and future research directions are provided in Sect. 5.

2 Related work

In 1992, Alan Mackworth published an article, 'On Seeing Robots.' In this article, he proposed that the robot football game is a good experimental platform for the study of artificial intelligence and robotics (Luo 2008). This view of Alan Mackworth has been unanimously endorsed by researchers because artificial intelligence can be better studied and implemented in soccer robot games. Robot football is a high-tech confrontation based on sports competitions. Playing football games in a dynamic and uncertain environment is a test of artificial intelligence. The research of robotic soccer involves multiple disciplines, including robotics, computer technology, electromechanical technology, intelligent control and decision-making, and sensor communication technology. In 1997, the first Robot World Cup football match was held in Nagoya, Japan, and achieved great success (Zhang 2005). Since then, robot football games have flourished and attracted more and more people's attention. This challenging work has also aroused intense interest among the research community.

In the context of robot football simulation, people's attention has been paid to artificial robot football matches in recent years. Many researchers have applied artificial intelligence algorithms on the simulation platform and achieved good results. The characteristic of the simulation game is that all hardware equipment is realized by computer simulation, which simplifies the complexity of the game system, reduces hardware requirements, has good controllability, nondestructiveness, reusability, and is not restricted by hardware conditions and venue environments. At the same time, it is also an excellent platform for studying artificial intelligence. The simulation robot football game is conducted on a standard computer software platform. Through this platform, participants can independently play the design action function, specify the strategy, etc. The process of football matches is more complicated. Refer to Peter Stone in Stone (1998), which divides the game process into multiple subtasks such as dribbling, passing, and shooting, and proposes a flexible multi-agent team structure. To adapt to the real-time and noisy environment, when designing the agent's basic actions and top-level decision functions, Peter Stone uses layer learning technology to decompose complex activities into different layers and complete them with other parts. Reference (Carvalho and Oliveira 2011) makes improvements based on the SARSA algorithm and puts forward a function approximation method CMAC (cerebellar model arithmetic computer); through the experimental results of ball control training, this method can effectively improve the dribbling of the agent. Effect. The literature (Stone et al. 2005) described the 3VS2 problem in detail. This article uses the SARSA algorithm and compares the experimental results of the agent in different grids. Passing strategy is an essential strategy during the game. It considers the long-term interests of the team. The reinforcement learning of passing decision-making is a typical reinforcement learning based on a competent team. The Q-Learning algorithm is applied in football robot simulation. Learning the passing strategy can achieve better results (Zhang et al. 2005). Reference (Iima and Kuroe 2008) improved the path planning strategy based on the SARSA algorithm and accelerated the agent's speed to obtain the maximum reward value. The Q-Learning algorithm can be used to learn the team's passing decisions effectively. On this basis, some scholars have proposed an algorithm that combines the Q-Learning algorithm with adversarial path planning (Jin et al. 2002). Good results have been achieved in playing football. In addition, to improve the team's defensive ability, they designed a complete defensive system (Yun et al. 2002). The robot football simulation platform can test new algorithms and new ideas and transform theoretical knowledge into practical applications to better promote the development of artificial intelligence.

3 Method

Machine learning algorithms play a significant role in almost every field of life. Figure 1 discusses the basics of machine learning process.

3.1 Reinforced learning model

In football and other sports, reinforced learning is gaining significant attention. This model considers the football ground and modeled it as a random infinite state machine with inputs and outputs.

State transition formula:

$$f_s = P\left(X_{(t)} | X_{(t-1)}, A(t)\right) \tag{1}$$

Observation value function:

$$f_o = P\left(Y_{(t)} | X_{(t)}, A(t)\right) \tag{2}$$

Reward function:

$$f_R = E\left(R_{(t)} | X_{(t)}, A(t)\right) \tag{3}$$

The entity is also modeled as a random infinite state machine with inputs and outputs.

State transition formula:

$$S(t) = f\left(S_{(t-1)}, Y_{(t)}, R_{(t)}, A_{(t)}\right) \tag{4}$$

Strategy/output function:

$$A_{(t)} = pi\left(S_{(t)}\right) \tag{5}$$

where the parameter X is the environmental value, A is the reward value, S is the strategic value, and Y is the value function.

The basic model is shown in Fig. 2. The goal of the entity is to learn a strategy to maximize the reward.

3.2 Several classic algorithms for reinforcement learning

3.2.1 Temporal difference learning (TD)

The most straightforward TD algorithm is the TD (0) algorithm. The TD (0) algorithm is an iterative strategy algorithm with adaptive capabilities. The update formula of the TD (0) algorithm value function is as follows:

$$V(S_t) \leftarrow V(S_t) + \alpha[r_{t+1} + \gamma V(S_{t+1}) - V(S_t)] \tag{6}$$

where α represents the learning factor, γ represents the discount rate, $V(S_t$ represents the value function of the entity at time t, state S_t, $V(S_{t+1})$ represents the state value function predicted by the entity at time $t+1$, state S_{t+1}, and r_{t+1} represents the entity from the state after the execution of the behavior, the instantaneous reward value obtained after S_t transitions to the S_{t+1} state.

3.2.2 Q-learning algorithm

The main task of the Q-Learning algorithm is to calculate the strategy π when the initial conditions are unknown. The

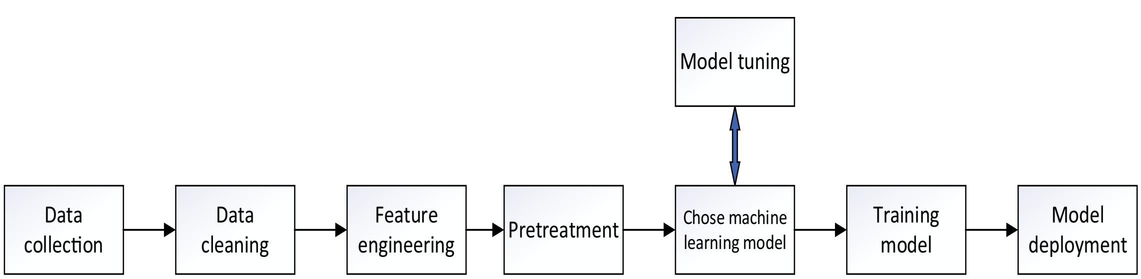


Fig. 1 Machine learning process

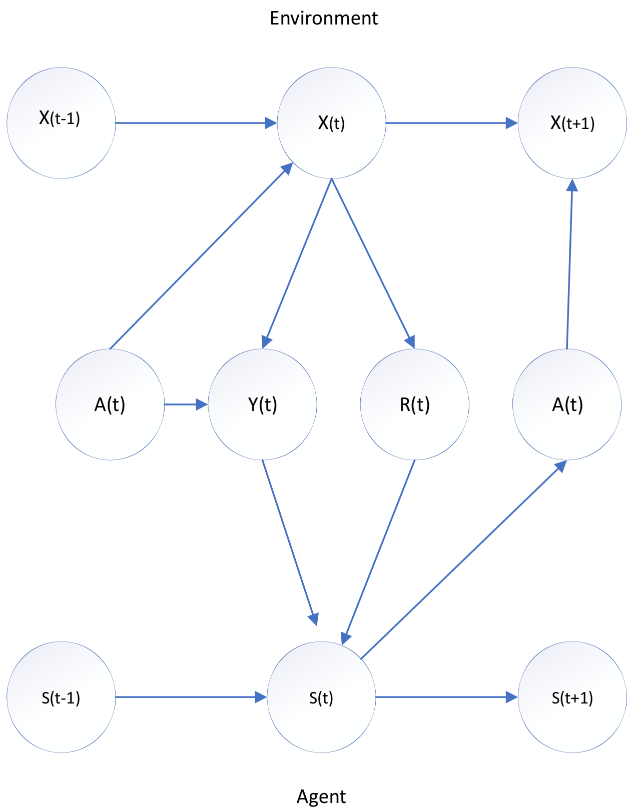


Fig. 2 Basic model of reinforcement learning

main idea of the algorithm is: Do not predict the environment model, directly optimize a Q function that can be calculated iteratively, and specify the reward function for the Q function when the Q function performing the behavior a_t in the state s_t, and then the enhanced signal discount when performing the optimal behavior with.

$$Q(s_t, a_t) = r_t + \gamma max Q(S_{t+1}, a) \tag{7}$$

Formula (7) is only valid when the optimal strategy has been obtained. In the early stage of learning, the observation function update rule of the Q-Learning algorithm is:

$$Q(s_t, a_t) = (1 - \alpha_t)Q_{t-1}(s_t, a_t) + \alpha_t[r_t + \gamma max Q_{t-1}(S_{t+1}, a)] \tag{8}$$

3.2.3 SARSA algorithm

The SARSA algorithm is a model-based reinforcement learning algorithm proposed by Rey and Niranjan in 1994. Because it still uses Q-value iteration, the algorithm was called an improved version based on the Q-Learning algorithm at the beginning. The SARSA algorithm is an online strategy Q-Learning algorithm. The behavior and environmental state values are estimated by the behavior taken by the current entity. The update of the Q-value in the SARSA algorithm can be expressed by the following formula:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \tag{9}$$

In each learning iteration, the entity first selects the behavior determined by the function according to the strategy and gets feedback, then chooses the behavior a_{t+1} when the process determines the next state s_{t+1} and modifies the observation function according to the formula (9). Then take the determined behavior a_{t+1} as the behavior of the next moment state; that is to say, the next moment behavior of the entity in the algorithm is determined by the current Q-value.

3.2.4 Optimize SARSA algorithm

The algorithm is based on the SARSA learning algorithm. The basic idea is to share a triplet such as a reward value r obtained by the entity performing a specific behavior in a particular state s to improve the updated degree of the entity's observation value function. This way improves the learning efficiency of entities. This improved version of the SARSA learning algorithm is generally used in multi-entity tasks.

4 Experiment and discussion

4.1 Discretization of simulation environment and basic action settings

(1) Discretization of the site: The simulation platform provides a complex continuous state environment. To better apply reinforcement learning, we first discretize the continuous state of the domain. In the MSRS platform, the game venue is 400 cm•280 cm in size. We divide it into 1602 areas. Except for the two goals, each area is 10 cm•7 cm in size. After 273 dividing the area, we use two state variables s_1, and s_2 to indicate the areas where the agent and the football may be located, so there are a total of 1602 •1602 = 2,566,404 states, as shown in Fig. 3. During the experiment, the target agent can be placed in any field area, and the football's location can be obtained through a written program. When the agent and the football are in the same area, we set the state to reach the goal.

(2) The design of the basic actions of the agent in a discrete environment: In a discrete state climate, there are five choices for the agent's actions: stay in

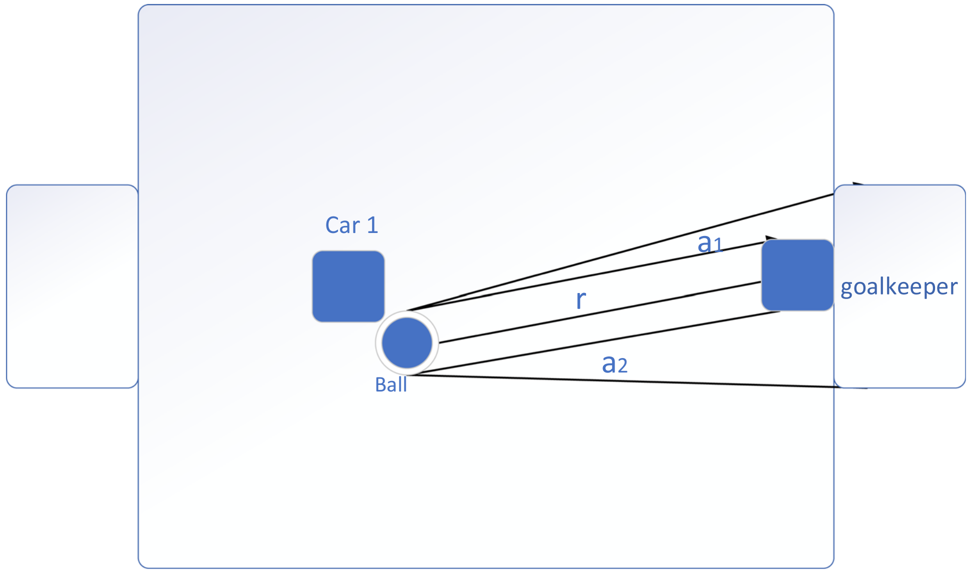


Fig. 3 State variables of the agent after gaining possession of the ball

place, move up one square, move down one square, move left one square, and move right one square.

4.2 Simulation experiment data analysis

4.2.1 Environment variable value setting

In the single-agent confrontation, the state of the agent when it does not control the ball is the same as in the single-agent experiment, so the simulation environment is discretized in the same way as in the single-agent experiment. When the agent gains possession of the ball, the state variables of the environment change; Fig. 3 shows that two angles, 'a_1' and 'a_2,' are formed between the ball and the opponent's goalkeeper and the two goalposts. The distance between the ball and the goalkeeper is 'r.' Then the state variables of the environment are determined by the values in 'a_1' and 'a_2.' The maximum angle and the distance 'r' are combined. Suppose state variables 's_1' and 's_2,' 's_1' represents the maximum angle a_{max} in 'a_1' and 'a_2,' and 's_2' represents the distance 'r.' The values of 'a_{max}' and 'r' are set as shown in Tables 1 and 2.

4.2.2 Looking for the ball

It can be seen from the results of the simulation experiment that the single agent learns slowly through continuous interaction with the external environment. This process can be clearly seen from the running track of the single agent from the initial state to the target state. After multiple runs, the trajectory of the agent is getting shorter and shorter.

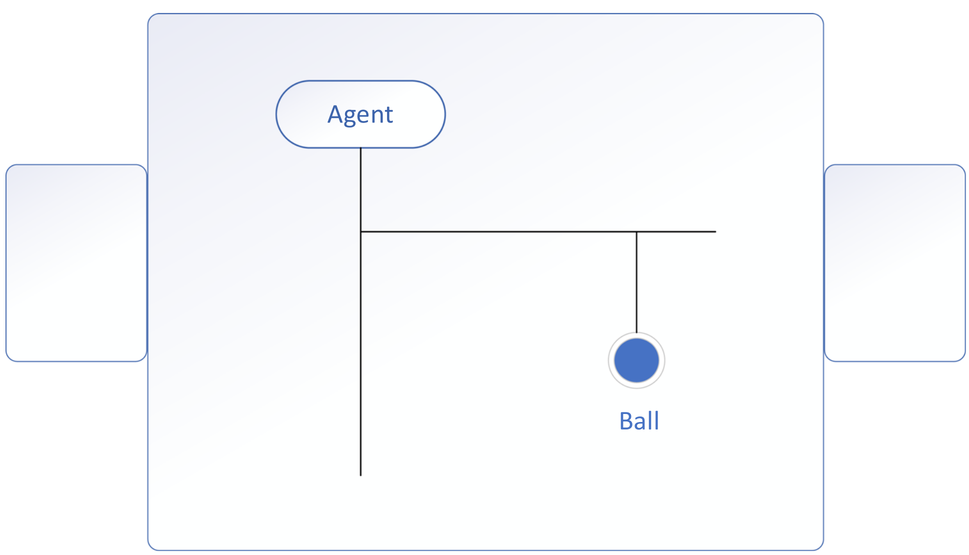


Fig. 4 The running trajectory of the agent reaches the target state during the first run

Table 1 The angle changes the specific setting of the state variable

Maximum angle of a_1 and a_2	State variables
$a_{max} < 15°$	$s_1 = 0$
$15° \le a_{max} < 30°$	$s_1 = 1$
$30° \le a_{max} < 45°$	$s_1 = 2$
$45° \le a_{max} < 60°$	$s_1 = 3$
$60° \le a_{max} < 75°$	$s_1 = 4$
$75° \le a_{max} < 90°$	$s_1 = 5$

Table 2 The specific setting of the distance changes the state variable

Distance between the ball and the goalkeeper	State variables
r < 40	$s_1 = 0$
$40 \le r < 80$	$s_1 = 1$
$80 \le r < 120$	$s_1 = 2$
$120 \le r < 160$	$s_1 = 3$
$160 \le r < 200$	$s_1 = 4$
$200 \le r < 240$	$s_1 = 5$
$240 \le r < 300$	$s_1 = 6$
$300 \le r < 400$	$s_1 = 7$
$400 \le r$	$s_1 = 8$

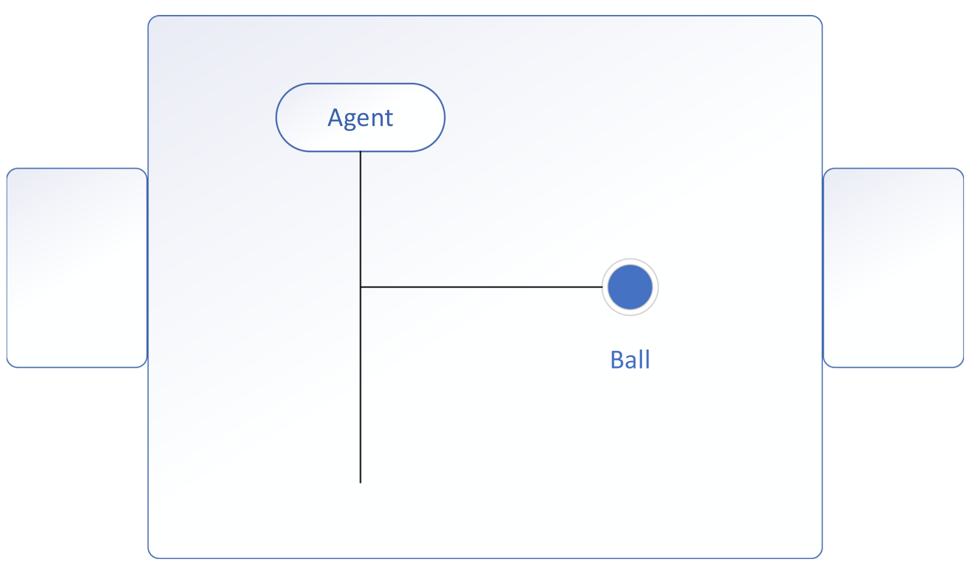
Figures 4, 5, 6 show that the trajectory of the agent is a broken line. This is because in this experiment, the agent can only move to the adjacent upper, lower, left, and right areas when selecting actions and cannot move along any angle. As can be seen from the above figure, the path of the agent running is relatively long at the beginning of the operation, and through the learning process of multiple 'trial and error,' the path gradually becomes shorter and finally reaches the optimum. Analyzing the process of the experiment, we found that for this relatively simple task of finding a ball, the agent can complete and find a more ideal path in a few learning cycles. Part of the reason is that the environment is relatively simple and there are no obstacles.

4.2.3 The influence of area divisions on the running path of the agent

This paper divides the competition field into 402 and 1602 areas to conduct an experiment of finding the ball by the agent. In the game mode, in the comparison of the two experimental results, series 2 represents the number of steps taken by the agent to reach the target state when the site is divided into 402 areas, and series 3 represents the number of steps taken by the agent to reach the target state when the site is divided into 1602 areas. The algorithm converges faster and can reach a better solution. Figure 7 briefly shows the convergence times of the agent in the learning process.

This Fig. 8 mainly studies various reinforcement learning algorithm application experiments on the robot soccer simulation platform and analyzes the experimental results. (1) The application of the improved SARSA algorithm in single-agent ball finding is studied, and the algorithm is applied to single-agent countermeasures. Using a hybrid greedy algorithm to update the reward value, it is found that good results can be achieved in fewer learning cycles. (2) The basic action design of the multi-agent, the role of the agent and the dynamic adjustment mechanism, and the reward value adjustment mechanism are studied. Through simulation experiments, it is found that the improved SARSA learning algorithm can better improve the ball

Fig. 5 Running trajectory of the agent reaching the target state during the 10th run



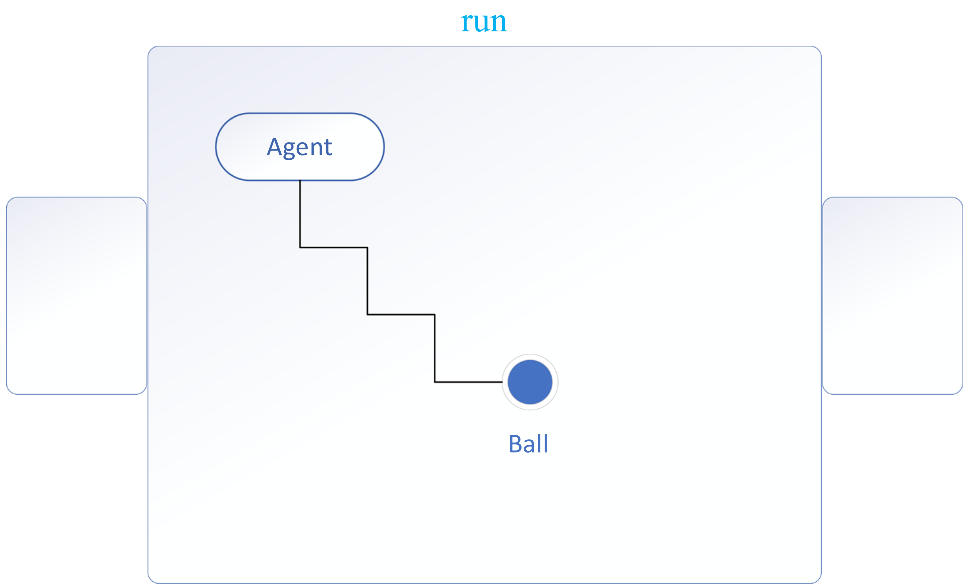


Fig. 6 Running trajectory of the agent reaching the target state during the 20th run

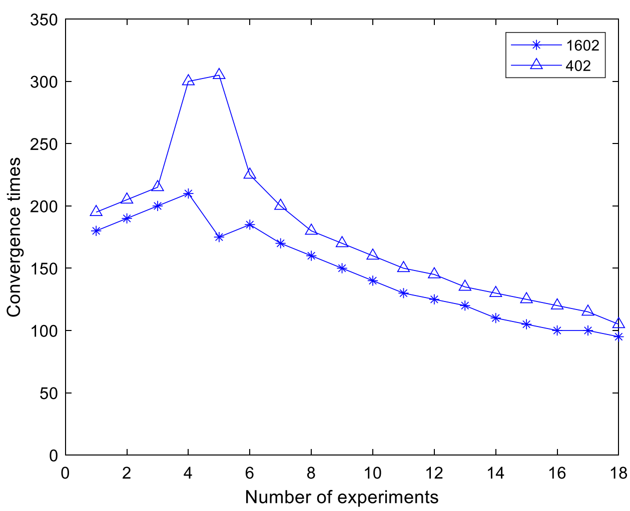


Fig. 7 Convergence times of the agent in the learning process

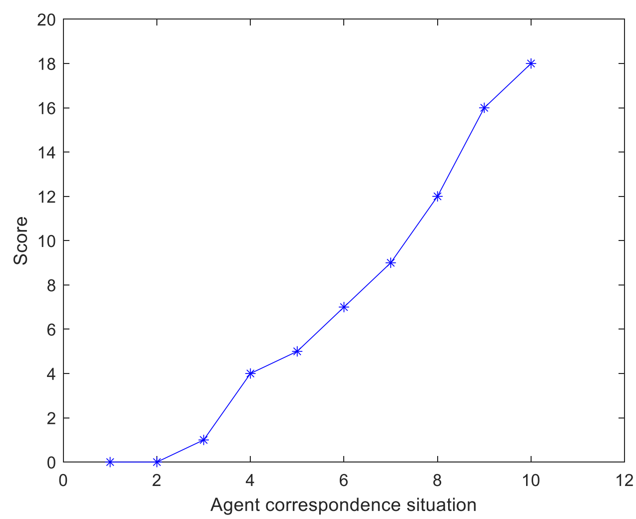


Fig. 8 Agent correspondence situation

possession ratio and steal ratio in multi-agent confrontation.

5 Conclusion

With the development and improvement of artificial intelligence, the application of robotic has felt its presence in various fields. On such application is the robot soccer game: a high-tech competition that integrates advanced. The main work of this paper is to apply an enhanced learning algorithm to the MSRS robot football game simulation platform. First, the basic principles of reinforcement learning are analyzed, and the Q-Learning algorithm, the instantaneous difference algorithm, and the SARSA algorithm are emphatically studied. On this basis, combined with the characteristics of robot soccer games, an improved/optimized SARSA algorithm is proposed. The optimized SARSA is a reinforcement learning algorithm that uses Q-value for making critical decisions about various parameters involved in the football training. Next, based on the simulation software platform of the robot football game, the reward function and behavior function are designed for the reinforcement learning algorithm in the simulation, and the behavior selection method and state discretization are discussed. Finally, the optimized SARSA algorithm is simulated and analyzed in a single-entity environment of a robot football game. On this basis, the application performance of SARSA and the optimized SARSA algorithm in a multi-entity environment are compared. The results show that the proposed approach outperforms the SARSA algorithm. The use of a flexible approach means that the study conducted in this paper can be applied to various other sport events.

Funding The paper did not receive any financial support.

Data Availability The data used to support the finding are cited within the article.

Declarations

Conflict of interest The authors declared that they have no conflicts of interest to this work.

Ethical approval This paper does not deal with any ethical problems.

Informed consent We declare that all authors have informed consent.

References

Carvalho A, Oliveira R (2011) Reinforcement learning for the soccer dribbling task. In: Proc of 2011 IEEE Conf on Computational Intelligence and Games. Seoul, IEEE Society, pp 95–101

Dewey D, Russell S et al (2015) Research priorities for robust and beneficial artificial intelligence. Ai Mag Artif Intell 47(2): 221–234

Dimeas AL, Hatziargyriou ND (2005) Operation of a multiagent system for microgrid control. IEEE Trans Power Syst 20(3):1447–1455

Elliott C (1992) The affective reasoner: a process model of emotions in a multi-agent system. Phd Thesis Northwestern University

Iima H, Kuroe Y (2008) Swarm reinforcement learning algorithms based on Sarsa method. In: Proc of the SICE Annual Conference. Tokyo, IEEE Society, pp 20–22

Jin YY, Jiang C, Zeng QS (2002) An application in RoboCup combining Q-learning with adversarial planning. In: Proc of the 4th World Congress on Intelligent Control and Automation. Berlin, IEEE Society. pp 496–500

Le QB, Park SJ, Vlek P et al (2008) Land-Use Dynamic Simulator (LUDAS): a multi-agent system model for simulating spatio-temporal dynamics of coupled human–landscape system I. Structure and theoretical specification. Ecol Inf 3(2):135–153

Lu Z, Li Y (1999) Interactive support vector machine learning algorithm and its application. J Tsinghua Univ 01995:334–355

Luo JT (2008) Research on decision-making system in MiroSot robot soccer. Central China Normal University, Wuhan, pp 1–6

Marjanovi M, KovaEvi M, Ba Jat B et al (2011) Landslide susceptibility assessment using SVM machine learning algorithm. Eng Geol 123(3):225–234

Stone P (1998) Layered learning in multi-agent systems. School of Computer Science Carnegie Mellon University, Pittsburgh, PA, pp 15213–23891

Stone P, Richard S, Sutton GK (2005) Reinforcement learning for Robocup Soccer keepaway. Int Soc Adapt Beh 13(3):165–188

Vidyasagar M (2015) Statistical learning theory and randomized algorithms for control. Control Syst IEEE 18(6):69–85

Wiley EO, Mcnyset K, Peterson T et al (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16(3):120–127

Xu XH (2004) Discussion on the military significance of robot football. Robot Technol Appl 10(2):8–9

Yun PC, Chen J, Yao JY et al (2002) Global planning from local eyeshot: an implementation of observation-based plan corrdination in Robocup simulation games. Robot soccer world cup V. Springer Verlag, Berlin, pp 230–255

Zhang J (2005) Design and research of soccer robot car controller based on embedded system. Central South University, Changsha, pp 10–17

Zhang JW, Hang GS, Zhang W (2005) Application of Q-learning algorithm in dribbling ball training of Robo cup. Syst Simul Technol 1(2):84–87

Zhao KQ (2007) The application of SPA-based identical-discrepancycontrary system theory in artificial intelligence research. ACM Transact Intell Syst 87(11):11–30

Zheng H, Zhou X, Duyan BI (2003) Introduction statistical learning theory and support vector machines. Mod Electr Tech 112(4): 5895–5960