Interactive teaching mode based on deep reinforcement learning

Jun Shao
Suzhou Industrial Park Vocational and Technical College, Suzhou, Jiangsu, 215000

Abstract: interactive teaching mode has been widely used in the teaching process of all levels and courses. With the deepening of interactive teaching, the restrictions on interactive time, interactive scene, interactive means and other aspects are becoming more and more obvious, and the current teaching needs can not be met only by classroom, students, pen and paper, spreadsheet, and slide teaching. This paper proposes a teaching framework based on deep reinforcement learning, which comprehensively enhances interactive teaching by integrating agents, simulation environment and data dashboard, so that interaction is no longer limited to interaction time, interaction space and interaction object, and then improves the usability and applicability of interactive teaching.

Key words: deep reinforcement learning; reciprocal teaching

1. Introduction

Since the birth of alphago in 2016, deep reinforcement learning has attracted much attention and entered the public view. Deep reinforcement learning is mainly used to solve sequential decision-making problems, such as: making the next action decision at every moment in the game; In supply chain management, purchase decisions are made at each time point (every day or every cycle). In these scenarios, it is necessary to select the behavior according to the current or historical data (state), and make the decision of the next behavior according to the feedback of the environment (change of state). In deep reinforcement learning, agents gain experience through continuous interaction with the environment, and learn from experience which behavior to take in different states to obtain the maximum cumulative benefit. Deep reinforcement learning is not only applied in the field of game AI, but also has been applied in operations research fields such as delivery lists, logistics path planning, online packing and so on.

On January 6, 2021, Ke Jie, the world go champion, said in an interview: “now they are basically training with AI, but rarely with people.” It can be seen that AI system based on deep reinforcement learning can be applied in the field of teaching or training. Taking supply chain management as a case, this paper discusses the application of deep reinforcement learning in teaching mode. In the teaching process of supply chain management, the scene of simulated operation is usually introduced to let students experience the process, methods and skills of supply chain management. Among them, “beer game” is a famous simulated “game”.

How does deep reinforcement learning work in such simulated “games”?Simulation “games” with cooperation or competition need an excellent collaborator and competitor. An excellent collaborator can improve students’ confidence and is also the object of study and research for beginners; An excellent competitor is not only the goal of students, but also the benchmark to test the level of students. In reality, whether students’ peers or opponents are excellent is a matter of luck or probability. Deep reinforcement learning can train one or more agents as collaborators or competitors. These agents learn by interacting with a large number of simulated environments, and their ability (performance) is tested and verified. Students at different learning stages can even choose different levels of agents to participate in the simulation. That is, deep reinforcement learning provides different levels of collaborators and competitors.

This paper proposes a framework that combines the deep reinforcement learning model with the interactive teaching mode. It solves the problem of unbalanced participants in the interactive teaching mode, and provides a theoretical basis and technical framework for the further development of interactive teaching.

2. Deep reinforcement learning

Reinforcement learning is an algorithm that trains agents to learn how to make decisions to maximize cumulative benefits / rewards through a large number of empirical data generated by interaction with the environment. In reinforcement learning, agents interact with the environmental system, that is, they choose a certain behavior to affect the environmental system and change the environmental data; The agent obtains the benefits of this behavior and its visible state data from the environment.

The core of reinforcement learning is how to determine the next behavior through state data and rewards, that is, learning a function to map the state and rewards to behavior, and deep reinforcement learning uses deep neural network to learn the mapping function. Different algorithms design different mapping functions and neural network structures to complete agent learning.

3. Interactive teaching

Interactive teaching was first proposed by palinscar in 1982 and has been applied by teachers all over the world. Interactive teaching is a process of interaction, influence and exchange between teachers and learners in the teaching process. Integrated teaching refers to the integration of theory and practice, teaching learning, learning to do, teaching to do.

Interactive teaching has achieved remarkable results in practice. Compared with traditional lecture based classroom teaching, interactive teaching not only provides students with opportunities to practice, but also increases the communication between teachers and students, but it also has some problems:

1) The teaching effect largely depends on the enthusiasm of the participants. If some participants do not cooperate, are not active or

have a large level gap, the effect will be greatly reduced, resulting in the reduction of other students' income.

2) It is impossible to guarantee the amount of practice of individual students. Although teachers carefully design the interactive process, the interaction is often unsustainable or repetitive, resulting in insufficient individual training, which is difficult to play the real effect of interactive teaching.

3) The role is relatively fixed. After the interaction starts, once the participants' roles are determined, their practice behavior will be bound to the roles, resulting in the practice training behavior will deviate in a fixed direction.

4) It is difficult to achieve ladder. The so-called ladder refers to the design of different interactive processes for students at different levels. In practice, on the one hand, it is difficult to actively assess the level of each student, on the other hand, it is difficult to meet the needs of each student with time and energy.

4. Case study

This paper uses the classic case of Supply Chain Management: "beer game" as the research object. The "beer game" simulates the inventory optimization scenario in supply chain management, in which four roles are set: manufacturer, distributor, wholesaler and retailer, and four participants play these roles respectively. In each round, retailers will receive customers' commodity demands, which can be generated randomly or designated by teachers. Retailers buy goods from wholesalers, wholesalers buy goods from dealers, dealers buy goods from manufacturers, and manufacturers produce goods themselves.

The inventory level of each role determines its cost. The inventory level is calculated as follows:

Inventory level = existing inventory + goods arrived in the current period – cumulative delayed delivery goods – goods delivered in the current period (equation 1)

When the inventory level is positive, each unit of goods will generate inventory cost; when the inventory level is negative, each unit will generate deferred cost.

5. Integrating deep reinforcement learning and interactive teaching

In interactive teaching, the following three basic modes can be summarized: interaction between students, interaction between teachers and students, and interaction between students themselves. Students can be collaborators or competitors. They all need to make decisions according to the current observed situation. These decisions will affect each other and lead to certain results. The relationship between teachers and students is generally that of adjudicators and participants. Adjudicators need to ensure that true and sufficient information is transmitted to all participants. Students' interaction with themselves refers to updating future strategies by observing historical data, which is a one-way interaction.

This paper proposes a framework of teaching mode based on deep reinforcement learning, which integrates the three modes

1) Agent convergence

Deep reinforcement learning algorithm can train agents to play a role in decision-making. Agents search for the best decision-making scheme by constantly trying and obtaining positive or negative rewards in the simulated environment. By setting reasonable rewards, they can guide agents to make correct decisions, and then train different levels of agents, who can obtain different amounts of rewards. In the "beer game", agents can play any role, and agents of different levels can control the total cost at different levels.

2) Simulated environment

Deep reinforcement learning needs to establish a simulation environment for agent training. The simulation environment here realizes the complete interactive rules, which can replace the role of the teacher as the adjudicator. As the adjudicator, the simulation environment can ensure the correctness of information transmission, and can provide adjudication services for multiple simulation scenarios at the same time. With the help of cloud platform, database, data visualization and other technologies, the simulation environment can provide a more intuitive, convenient and easy-to-use interface environment.

3) Data dashboard

The simulation environment will automatically record the interactive data, which will be provided to the agent for training in the early stage. The statistics and summary of these data into a visual data dashboard can be used as a powerful basis for data analysis. The data dashboard provides historical data statistics of each bureau simulation, which is convenient for students to learn from experience. It is an interesting coincidence that agents also analyze and learn from these historical data. The difference is that the agent is guided by the rewards set in the simulation link; Students learn through statistical analysis tools and experience summary.

6. Experiment

In this paper, the PPO algorithm based on deep reinforcement learning is used to train the agents of four roles in the beer game. The human strategy is used to simulate in the same simulation environment, and the average total cost of 100 rounds is compared between the two.

Cases 1-6 use different random models to generate retailer customer demand. The first three use uniform distribution and the last three use normal distribution, and their variance also increases. Experiments show that the cost control of agents is lower than that of human strategies in all cases, especially when the variance is greater, that is, the demand changes and fluctuations are greater, the difference between the two cost control is greater.

Many universities and enterprises have developed the “beer game” simulation environment and configured the corresponding data dashboard. Most simulation environments provide a simple and easy-to-use graphical interface, and the entry level of operation is almost zero; The visualization of various data indicators in the simulation process by the data dashboard provides an intuitive tool and basis for students’ subsequent analysis.

7. Summary

In interactive teaching, the integration of agent, simulation environment and data dashboard enhances the three interactive modes in the process of interactive teaching, which makes interactive teaching adapt to different levels of students; It liberates teachers’ repetitive and mechanical responsibilities, so that teachers can pay more attention to students and the interaction process; It makes the interaction no longer limited to time and space.

References

[1] By Chunxiu Pei and Chunying Yang Research on interactive teaching mode [j]Journal of Hunan University of science and technology, 2007, issue 05 (5): 137-139

[2] Written by Guiling Zhao Application of traditional interactive teaching mode in network environment [j]Experimental science and technology, 2009, issue 1: 114-117

[3] Bowen, W.G., chingos, M.M., lake, K.A., & Nygren, T.I. interactive learning online at public universities: evidence from randomized trials[j]Ithaka, may 22, 2012:10-25

[4] A. oroojlooyjadid, M. Nazari, L.v. Snyder, M. Tak á C A deep q-network for the beer game: deep reinforcement learning for inventory optimization[j] Manufacturing & service operations management, 2021:5-15

[5] R. S. Sutton and A. g. bartoReinforcement learning: an introduction[m]MIT Press, Cambridge, 1998:156-178

[6] Y. liDeep reinforcement learning: an overview[j]ArXiv preprint arxiv:1701.07274, 2017:15-23

[7] Schulman, John, et al. proximal policy optimization algorithms[j]ArXiv preprint arxiv:1707.06347, 2017:2-5

[8] J. D. stermanModeling managerial behavior: deviations of feedback in a dynamic decision making experience[j]Management science, 35 (3), 1989:321 – 339

[9] Kaelbling, Leslie P; Littman, Michael L; Moore, Andrew W. reinforcement learning: a survey[j]Journal of artistic intelligence research4, 1996: 237 – 285

[10] Bellman, R. a Markovian decision process[j]Journal of mathematics and mechanics6 (5), 1957: 679 – 684

[11] H. L. Lee, V. Padmanabhan, and s. whangInformation disturbance in a supply chain: the bullwhip effect[j]Management science, 43 (4), 1997:546 – 558

[12] I. giannoccaro and P. pontrandolfoInventory management in supply chains: a reinforcement learning approach[j]International Journal of production economics, 78 (2), 2002:153 – 161

[13] S. K. chaharsooghi, J. heydari, and S. H. zegordiA reinforcement learning model for supply chain ordering management: an application to the beer game[j]Decision support systems, 45 (4), 2008:949 – 959

[14] Ben David, Shai, kushilevitz, Eyal, Mansour, yishayOnline learning versus offline learning[j]Machine learning29 (1), 1997: 45 – 63

[15] Richard suttonLearning to predict by the methods of temporary differences[j]Machine learning3 (1), 1998: 9 – 44

This paper is the result of the 14th five year plan of Jiangsu Educational Science in 2021. The project name is the research on the integrated interactive teaching mode based on deep reinforcement learning (project number: 2022sjk02rz01).