

Analysis of the Evolution and Causes of Educational Involution Based on Prisoner's Dilemma and Reinforcement Learning

Yicheng Gong, Yanli Xu*, Qing Liu, Yuqiang Feng

Faculty of Science, Wuhan University of Science and Technology, Wuhan, Hubei, China

Abstract

The severe situation of adolescent growth is often attributed to vicious competition leading to educational involution. In this paper, the frequency of "focus on scores" is increased and the return is significantly reduced as a sign of educational involution, and an educational game is constructed to analyze the choice and balance of "focus on scores" and "focus on happiness" for "home-school students". Because education is not easy to experiment, in order to reveal the evolution process of educational games in reality, Q-learning was used to conduct 10,000 simulations. The results show that in the early stage, the frequency of "focus on scores" increased slowly and was lower than 50%, but the return did not decrease significantly and involution did not form; in the mid-term, the frequency of "focus on scores" increased rapidly and the return of those with higher frequency decreased significantly, and involution was formed; in the later stage, the return of those who entered the involution first dropped to the bottom, but then the frequency of involution and returns are overtaken and slowly rising; finally, the frequency of "focus on scores" converges at 70% to 82%, and involution is deadlocked. Therefore, in order to avoid educational involution, the frequency of "focus on scores" is preferably between 35% and 45%.

Key Words:

youth education; involution; Prisoner's Dilemma; Reinforcement Learning; Q-learning

1. Introduction

With the continuous development of higher education in China, youth education has been greatly popularized. In 2020, China's nine-year compulsory education rate has been consolidated at 95.2%, and the gross enrollment rate of higher education has reached 54.4%^[1]. However, in recent years, there have been many problems in adolescent growth, among which mental health problems have become very serious: 《Report on National Mental Health Development in China (2019-2020)》 shows that 24.6% of adolescents are depressed, of which 7.4% are severely depressed^[2]. This problem has attracted widespread attention in the society, and education has become an important breakthrough in improving the mental health of adolescents: on the one hand, because most of them are students, and the students spend about 40 weeks in school each year; on the other hand, the vicious competition of educational subjects has become the reason for the high exposure rate of children's psychological problems. Vicious competition is mainly reflected in the high academic pressure, and the high academic pressure is mainly due to the dual pressure of the amount of homework and the difficulty of homework for young people. The phenomenon of vicious competition in education is similar to involution: under the premise that the edge is fixed, it turns to the pursuit of internal refined development due to the depletion of creativity^{[3][4]}. To improve the problems of adolescent growth, we can start from education, clarify the evolutionary laws and causes of educational involution, and then seek solutions. At present, many experts and scholars have analyzed the causes of education involution from the four aspects of society, government,

AHPCAI2022@2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence
EMAIL: email: gongyicheng@wust.edu.cn (Yicheng Gong), * Corresponding author: xuyanli@wust.edu.cn (Yanli Xu), email: liuqing@wust.edu.cn (Qing Liu), email: yqfeng6@126.com (Yuqiang Feng)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

family and school. They believe that the social reasons are mainly due to the large gap between the rich and the poor, class solidification, and uneven distribution of social resources, etc.; the main reasons for the government are the single education evaluation system, the problem of high school entrance examinations, and the uneven distribution of educational resources, etc.; the main reasons for schools are limited teachers and problems with educational concepts, etc.; the main reasons for the family are the high demand for education and the existence of problems in the concept of education, etc. [5-9]. Experts and scholars put forward solutions based on these reasons, but did not discuss the problems in the interaction of educational subject strategies that lead to the formation of educational involution.

The vicious competition of educational subjects is an important reason for the formation of educational involution, and vicious competition shows that there is a problem in the way of strategic interaction of educational subjects, breaking the balance between the two strategies of "focus on happiness" and "focus on scores", making "focus on happiness" gradually shift to "focus on scores", but individuals face the result that "return effort ratio" both decline. The educational dilemma caused by this educational game is the same as that of the prisoner's game [10]: individual rationality leads to collective irrationality. Therefore, this paper chooses the game perspective analysis, builds a two-player educational game based on the prisoner's game, and analyzes its theoretical equilibrium situation and the causes of involution. In order to test the rationality of the theoretical analysis in this paper, a game experiment is needed. Education is related to the growth of young people and the future of the country. It is not tolerated to do game experiments easily. At the same time, people in real life have limited rationality. Therefore, reinforcement learning is considered for simulation experiments. Reinforcement learning [11] has achieved excellent results in many practical applications, such as games, robot control, finance, medicine, resource optimization scheduling, Industrial Process Control et al. Therefore, this paper uses the Q-learning algorithm in reinforcement learning, and regards the players in the educational game as agents, so that they can learn by trial and error in a simulated educational environment, and conduct simulated experiments on the educational process. Taking the high frequency of "focus on scores" and decreasing returns as the main signs of the formation of educational involution, this paper analyzes the evolution and causes of Chinese youths' educational involution, and explores the optimal strategy choice of bounded rational people and self-balance of "focus on happiness" and "focus on scores", to develop recommendations for improving educational content.

2. Game Analysis of Educational Involution from the Perspective of Prisoner's Dilemma

The important reason for the formation of educational involution is the vicious competition of educational subjects. Vicious competition indicates that there is a problem in the way of strategic interaction of educational subjects. Therefore, the educational involution can be analyzed from the perspective of game theory. The result of the educational game is that the individual faces a decrease in the "return effort ratio", which is the same as the essence of the prisoner's game: individual rationality leads to collective irrationality, so the prisoner's dilemma model can be used to analyze the evolution and causes of educational involution. Since the educational game in reality is particularly complex, in order to simplify the analysis and make the educational game more reasonable, the following five model assumptions are made.

2.1 Model assumptions

Hypothesis 1: Combine the parents, schools and students in the education subject into a rational person "home-school students".

Hypothesis 2: The short-term goal of the "home-school students" in the game is to maximize the score.

Hypothesis 3: There are many rational players playing games with each other in the whole education. In order to simplify the complexity of the model, two players "Home School Student A" (hereinafter abbreviated as HSSA) and "Home School Student B" (hereinafter abbreviated as HSSB) are selected to construct an educational game.

Hypothesis 4: Considering that people gradually change from "focus on happiness" to "focus on scores" in the educational game, it is assumed that in the educational game, HSSA and HSSB have two strategic choices: "Focus on Happiness" (hereinafter abbreviated as FH) and "Focus on Score" (hereinafter abbreviated as FS).

Hypothesis 5: When two rational people choose FH at the same time, the return obtained is R; when they choose FS at the same time, the return obtained is P; when they choose FH and FS respectively, the return obtained by the rational person who chooses FH is S, and the return obtained by the rational person who chooses FS is T.

At the same time, the relationship between the four returns satisfies the following five inequalities:

Inequality 1: $R > P$, if both parties choose FH at the same time, the returns are higher than if both parties choose FS at the same time.

Inequality 2: $R > S$, if both parties choose to FH at the same time, it is more profitable than if the other party chooses to FS.

Inequality 3: $2R > S + T$, compared with only one party choosing FH, the returns of both parties choosing FH at the same time are higher.

Inequality 4: $T > R$, only one party chooses to FS, this party can obtain the highest returns.

Inequality 5: $P > S$, when one party chooses FS, the other party chooses to FS with higher returns than FH.

2.2 Model establishment and theoretical analysis

Based on the above assumptions, the two- player education game model is constructed as shown in Table 1.

Table1. Two-player education game model

HSSA	HSSB	
	FH	FS
FH	(R,R)	(S,T)
FS	(T,S)	(P,P)

Analysis of Table1 shows that in the single two-player education game model, the Nash equilibrium strategy of HSSA and HSSB is FS, and the return is P; but the strategy that achieves Pareto optimality is HSSA and HSSB choose FH, and the return obtained is R. In the single two-player education game model, the players in the game can only reach the Nash equilibrium and cannot obtain the maximum return. When people chose FS, it will lead to the formation of educational involution. The educational dilemma caused by this educational game is the same as that of the prisoner's game: individual rationality leads to collective irrationality. Therefore, the existing conclusions and research results of the prisoner's game can be used to assist the discussion of the educational game.

Real education is continuous, so it can be seen as a repeated educational game, similar to the repeated prisoner's game. Theoretically, when the game is played repeatedly, HSSA and HSSB will realize that chose FH will yield greater returns. At this time, chose FH may appear as a balanced result. After repeating the game model repeatedly and nearly infinitely, the Nash equilibrium will tend to be Pareto optimal, from choosing FS to choosing FH. On the other hand, if all choose FH completely and do not study at all, it is not conducive to rejuvenating the country through science and education and the development of the country. Therefore, balancing FH and FS is a key issue in education. In order to test the practical effect of theoretical analysis, the following will explore the evolution law of educational game through the Q-learning algorithm in reinforcement learning.

3. Exploration of Evolutionary Law of Educational Game Based on Reinforcement Learning

There are two reasons for choosing reinforcement learning to explore the evolution law of educational involution: (1) Education is related to the growth of adolescent e and the future of the country.

Therefore, game experiments cannot be done at will, but reinforcement learning can simulate the models described by game theory and conduct game experiments. (2) In the game model, it is assumed that the people in the game are completely rational, but in real life, people are usually bounded rationality, and people cannot understand the environment they are in and accurately predict the future, that is, all information is not completely certain. The agent in reinforcement learning learns through trial and error under uncertainty, and finally acquires a strategy that maximizes the agent's expected return. Based on this, in order to test the analysis results of this paper, the Q-learning algorithm in reinforcement learning is used to simulate the educational game process, and to explore the evolution law and causes of educational involution.

3.1 Experimental platform and experimental design

This paper uses the python3.8 compiler to implement the reinforcement learning algorithm with numpy and panda function modules. Data visualization using the matplotlib library. In the experimental design, HSSA and HSSB both use the Q-learning algorithm to learn the strategy. The learning rate α of the algorithm is set to 0.1, the discount factor $\gamma=0.6$, and the number of iterations is set to 10,000 rounds. The update formula of the state action of the Q-learning algorithm to the value function and the probability update formula of the action selection are as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right) \quad (1)$$

$$P(a_i) = \frac{e^{Q(s,a_i)/\lambda}}{\sum_{a \in A} e^{Q(s,a)/\lambda}} \quad (2)$$

α is the learning rate, the larger the α , the faster the Q value converges, but the easier it is to oscillate; γ is the discount factor, which indicates the degree of influence of future rewards on the current action. Next, the Q-learning algorithm in reinforcement learning is used to simulate the repeated two-player educational game model to explore the evolutionary law of educational involution.

3.2 Analysis of experimental results

In order to analyze how HSSA and HSSB choose FH and FS as the number of iterations increases, after 10,000 iterations of the two- player education game model, a line graph of cumulative times is drawn as shown in Figure1.

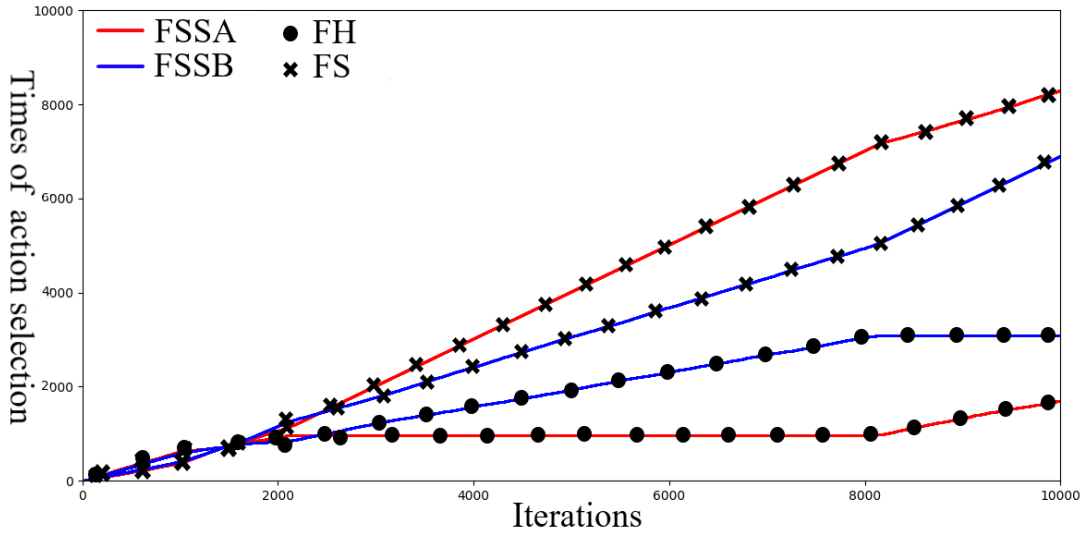


Figure1. Cumulative times of HSSA and HSSB choosing FH or FS

It can be seen from Figure 1: (1) With the increase of the number of iterations, both HSSA and HSSB gradually choose FS, indicating that in youth education, over time, people tend to choosing FS. (2) In

the vicinity of 1800 iterations, the accumulative times of the selection actions of HSSA and HSSB coincide, indicating that a demarcation point is entered at this time. Before 1800 rounds, the frequency of HSSA and HSSB choosing FH and FS was not much different, that is, there was no obvious preference for the two strategies at this time. After that, HSSA and HSSB gradually chose FS.

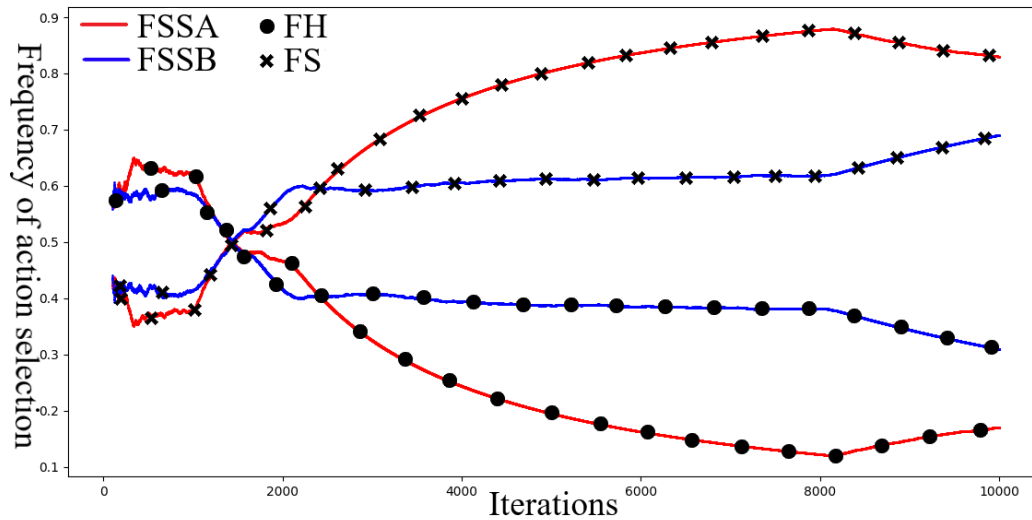


Figure 2. Frequency for HSSA and HSSB selection of FH or FS

It can be clearly seen from Figure 2: (1) Before the 1800 iterations, HSSA and HSSB chose FH more frequently, and HSSA chose FH more frequently than HSSB high. (2) After 2500 iterations, the frequency of HSSA choosing FS began to be gradually higher than that of HSSB; after 8100 rounds, HSSA began to reduce the frequency of selecting FS, while HSSB On the contrary, the frequency of HSSA choosing FS is still higher than that of HSSB, but there is a trend of convergence between the two.

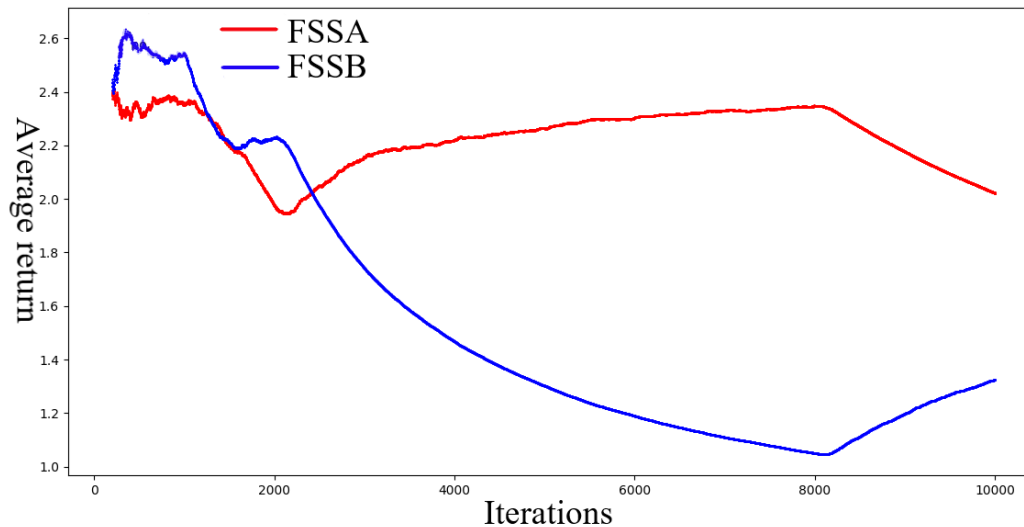


Figure 3. Average return of HSSA and HSSB

It can be seen from Figure 3: (1) The average return of HSSA and HSSB are the highest before 1800 rounds. At this time, the highest average returns of HSSA and HSSB are respectively are 2.00 and 2.65. Combining with Figure 2, it can be seen that at this time, HSSA and HSSB select FS at frequency of 34% and 42%, respectively. (2) The average return of HSSB gradually decreased from 2500 rounds to 8100 rounds, and was much lower than that of HSSA, while in the first 2500 rounds, the average return of HSSB was higher than that of HSSA. This phenomenon shows that people who take the lead in increasing the frequency of FS can make their own average returns higher than others in the short term, but after long-term games, their average returns will decrease rapidly and are much lower than others. Therefore, in youth education, those who take the lead in increasing the frequency of FS will be detrimental to others and themselves.

Comprehensive analysis of Figure 1, Figure 2 and Figure 3 can be obtained: (1) There is a process of involution from brewing to formation, then intense and then deadlocked. (2) The specific evolution law of educational involution is as follows: in the first 1800 rounds, the frequency of players choosing FS increased slowly and was lower than 50% but the average return did not drop significantly, and involution was in the brewing period. From the 1801st round to the 2500th round, the frequency of the two players choosing FS increased rapidly, and the average return of the player with higher frequency decreased significantly, and the player entered the involution first, and the involution was in the formation stage. From the 2501st round to the 8100th round, the frequency of the first invokers to choose FS reached about 60%, but the average return dropped significantly to the bottom, but then the frequency and average return of the invokers to choose FS surpassed and rose slowly; the involution was in an intense period. From 8101 rounds to 10000 rounds, the frequency of players choosing FS converged between 70% and 82%, and the average return of those who entered the involution first increased, while the average return of those who entered the involution decreased. A new round of involution is showing signs, but the frequency of the two choosing FS is very high, it is difficult to exert force again, and the involution is in a deadlock. (3) At the demarcation point formed by educational involution, the frequency of players choosing FS is 50%, and this also indicates that the natural balance of choosing FH and FS is about to be broken. In the later stage, although the frequency of players choosing FS decreased for a short time, it is still higher than 50%. Therefore, the frequency of home-school students choosing to focus on scores is higher than 50% as one of the main causes of educational involution. (4) In the stage where the educational involution is not yet formed, the average return of the players in the two games is the largest, which can be obtained by combining the three graphs. With a frequency of about 35% to 45%, the players in the game can get the best return, and possibly avoid the formation of educational involution.

4. Summary and Outlook

The main contributions of this paper are as follows: (1) From the perspective of the Prisoner's Dilemma, it theoretically analyzes the reasons for the formation of youth education involution; (2) The Q-learning algorithm in reinforcement learning is used to simulate a bounded rational player education game experiment; (3) Analyzed the evolutionary laws and causes of educational involution, and explored an optimal strategy: "focus on scores " with a frequency of about 35% to 45%.

Future work can be considered from the following aspects: (1) Differentiating the players in the model. When modeling, the two players in the education game are assumed to be indistinguishable people by default, but in reality, there are individual differences among the educational subjects. In the later stage, the individual differences of educational subjects can be specifically analyzed, and an asymmetric educational game model can be constructed. (2) To explore how long it will take to change the status of education involution after the country proposes the "double reduction" policy.

5. Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (72031009) and Hubei Province Key Laboratory of Systems Science in Metallurgical Process(Y202105).

6. References

- [1] China Youth Daily. Ministry of Education of the People's Republic of China: In 2020, the nine-year compulsory education consolidation rate was 95.2%, and the gross enrollment rate of higher education was 54.4% [DB/OL]. <https://baijiahao.baidu.com/s?id=1693014680894029763&wfr=spider&for=pc,2021-03-01 15:43>
- [2] Xiaolan Fu, Kan Zhang, Xuefeng Chen, etc. Report on National Mental Health Development in China (2019-2020)[M]. social sciences academic press(CHINA).2021:143-164.

- [3] Goldenweiser A. Loose ends of theory on the individual, pattern, and involution in primitive society[J]. *Essays in anthropology*, 1936: 99-104.
- [4] Hong Wang, Zhi Chen. On the Logic and the Path of “Double Reduction” Policy through the Perspective of Involution[J]. *Education&Economy*, 2021, 37(06): 38-43+61.
- [5] Zhujun Huang. On the Involution of Education in the Transition Period and Its Deciphering Path [J]. *Journal of East China Normal University (Educational Sciences)*, 2012, 30(02): 37-41+47.
- [6] Rong Mao. Universal Higher Education and Justice as Fairness[J]. *Jiangsu Higher Education*, 2021(08): 1-6.
- [7] Xiong Yang. The Roots and Cracks of "Educational Involution" in the AI Era [J]. *Social Sciences Digest*, 2021(11): 4-6.
- [8] Youhua Chen, Guo Miao. Enrollment Tournament, Educational Involution and Stratification of School District [J]. *Journal of Jiangsu Administration Institute*, 2021(03): 55-63.
- [9] Cheng Chen, Lei Bao. The Origin Involution and Solutions to Address Involution in Education [J]. *China Examinations*, 2022(02): 81-88.
- [10] Axelrod R, Hamilton W D. The evolution of cooperation[J]. *science*, 1981, 211(4489): 1390-1396.
- [11] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018: 1-22.