https://doi.org/10.26599/IJCS.2022.9100005 International Journal of Crowd Science

# Evolution of Agents in the Case of a Balanced Diet

Jianran Liu[1] and Wen Ji[1] ✉


## ABSTRACT

Agents are always in an interactive environment. With time, the intelligence of agents will be affected by the interactive environment. Agents need to coordinate the interaction with different environmental factors to achieve the optimal intelligence state. We consider an agent's interaction with the environment as an action-reward process. An agent balances the reward it receives by acting with various environmental factors. This paper refers to the concept of interaction between an agent and the environment in reinforcement learning and calculates the optimal mode of interaction between an agent and the environment. It aims to help agents maintain the best intelligence state as far as possible. For specific interaction scenarios, this paper takes food collocation as an example, the evolution process between an agent and the environment is constructed, and the advantages and disadvantages of the evolutionary environment are reflected by the evolution status of the agent. Our practical case study using dietary combinations demonstrates the feasibility of this interactive balance.



## KEYWORDS

intelligence evolution; agent interaction; agent-environment framework; universal intelligence


With the development of artificial intelligence technology, highly intelligent individuals have improved from being the original simple human or animal to individuals with high computing power. The general intelligent measurement method is usually used to evaluate the intelligence quantity of an agent. Human intelligence is measured by IQ tests[1] or cognitive ability questionnaires. Machine intelligence is usually measured by the Turing test[2,3] or other intelligence metrics[4]. However, the environment in which an agent is located is complex, and the agent cannot exist independently from its environment. Therefore, the measurement of agents' intelligence not only is related to their ability to complete a task[5], but also needs to consider the environment around agents. In their previous work, Ji et al.[6] put forward the intelligence evaluation of an agent using the quality-time-complexity group intelligence metric. They take into account the complexity of intelligence testing, rewards of the environment to the agent, and timeliness of the tests.

In general, the environment is composed of more than one factor. We draw on this idea so that the agent we design can subjectively assign the degrees of interaction between different factors in the environment, and the variables in the environment will give the agent corresponding feedback according to the agent's choice. Intelligence interacts with a variety of factors in the environment. Figure 1 shows how agents interact with the environment. Agents interact with other agents and with the environment. Each of these environmental factors includes a granular local interaction environment. For example, in the interactive environment of a school, there is a need to interact with different disciplines to enable the evolution of individual intelligence in all its aspects. In the interactive environment of diet, there are a variety of foods, and agents need to take nutrients from different foods to maintain their balance.

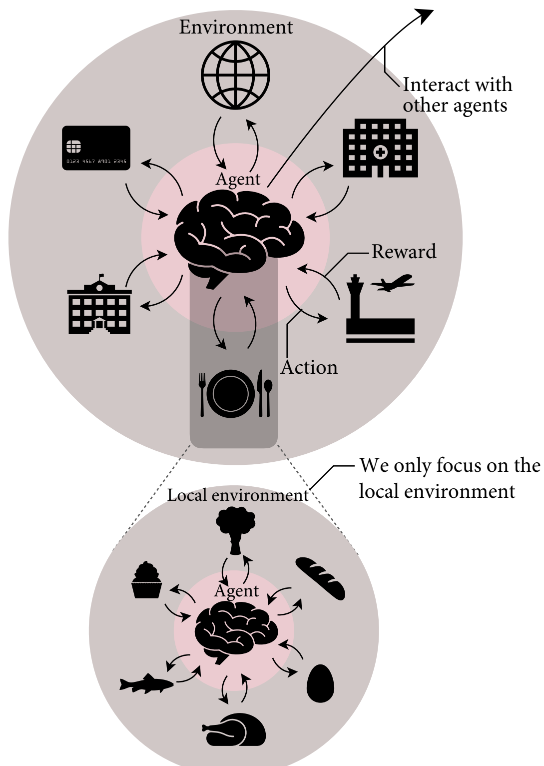


**Fig. 1 Interaction mode between the agent and environment. The agent needs to choose an optimal action strategy independently to make the sum of rewards of all environmental factors feedback reach the optimal state. The research object of this paper is the subenvironment of dietary collocation.**

In this paper, we take the dietary data as the reference environment, and focus on the evolution of agent in this environment. We take its health degree as its intelligence indicator, and simulate the intelligence evolution mechanism of agent under the influence of surrounding environment. First, we

1 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
Address correspondence to Wen Ji, jiwen@ict.ac.cn

constructed the principle of agent evolution and optimization objectives and locked the scope of optimal solutions. Then, we set parameters for each environmental factor, making the abstract problem concrete. Finally, we analyzed the sensitivity of the parameters that might cause fluctuations in the evolution. The results show that our proposed evolutionary framework can effectively describe the evolution of the real world. The contributions of this paper are summarized as follows:

● We take eating and drinking as intelligent evolution scenarios and treat the health status of an agent as an intelligence quantity.

● We consider the degree of interaction between an agent and an environmental factor as an exchange of equal information. When an agent acts in accordance with an environment, the environment will also provide feedback to the agent with its corresponding properties.

● We explore a new evolutionary model, which aims not to blindly seek the maximum intelligence of an agent but to achieve a balanced state through the appropriate interaction of different environmental factors.

The following sections are arranged as follows: Previous works are presented in Section 1. The proposed setting of evolutionary principles and instantiation parameters are explained in Section 2. The experimental verification of the examples set in Section 2 is discussed in Section 3. The proposed work is summarized and plans for future works are given in Section 4.

## 1 Related Works

### 1.1 Crowd intelligence and measurement

Crowd intelligence[7] is a new generation of intelligence. Its constituent form is no longer limited to all kinds of tangible entities from the physical world but also includes electronic intangible resources, such as software, data, and information, and even some intelligent resources that change with the state of agents. The cooperation between these heterogeneous resources and agents reflects the subversive transformation of the surrounding environment by agents and the influence of the environment on intelligent individuals[8].

The intelligence of an agent is a description of its ability and does not have a physical form itself. It can only reflect its value by relying on the carrying system. The value of intelligence is often reflected in the tasks performed, services provided, or evaluation indicators. In previous works, some scholars proposed to use intelligent entropy as a measurement model to measure the intelligence quantity and applied entropy theory to the analysis of functional variations.

Agents usually have a strong logical reasoning ability[9]. Therefore, agents in an interactive environment can cooperate with environmental factors to achieve the purpose of complementary advantages[10]. For example, various probabilistic models designed by the human brain can reach optimal conclusions using computing equipment and constantly adjust with new information, making processing systems specialized along with tasks to perform complex functions efficiently[11]. Furthermore, with the advanced analysis and learning models designed by the human brain, various possible results can be predicted. Such results determine the optimal cooperative working mode of the group system and reflect the autonomy of the heterogeneous group system in the evolutionary process.

However, the interaction between an agent and its surroundings is dynamic. In this process, an agent needs to measure its intelligent state in real-time, recognize the defects of its state, and then selectively continue to interact with the surrounding environment. In this paper, a model of the dynamic intelligent evolution is established.

### 1.2 Reinforcement learning and agent evolution

The idea of reinforcement learning[12,13] is based on the learning process of an agent. Here, an agent that is aware of the interactive environment needs to constantly interact with the environment to choose the best interaction means to evolve. The environment will make corresponding feedback and information output for each action of the agent. Therefore, reinforcement learning is a dynamic process. However, more often than not, an agent in reinforcement learning terminates reinforcement learning after finding an interactive strategy that maximizes its reward value[14]. Therefore, reinforcement learning in the traditional sense solves a convergent convex optimization problem.

The method proposed in this paper considers lessons from the interaction between agents and the environment in reinforcement learning. However, unlike reinforcement learning, we do not solve a convergent convex optimization problem but strive to make agents interact with the surrounding environment to form a dynamic balance. We use the dynamic stability of agents to indirectly evaluate whether the environment is conducive to intelligent evolution.

## 2 Construction of the Proposed Evolutionary Model

### 2.1 Principles of evolution

Unlike previous evolutions, our proposed evolution tends to maintain a balance, which can be interactive between an agent and its environment. Here, more interaction with the environment does not mean better for the agent. We need to select the best interaction scheme in the surrounding environment so that agents can be stable during the evolution.

For the interaction solution, two key points need to be determined:

● Specific environmental factors that interact with agents.

● The amount of interactive information between the agent and environmental factors.

During each evolutionary iteration, the sum of the interaction information between an agent and various environmental factors is fixed. Before the interaction, agents need to assign interactive information to all environmental factors. The information that can provide a reference for an agent includes the number of environmental factors, attributes of environmental factors, and the current state of the agent.

The distribution of the total amount of information interaction involves a quantitative problem. Because it is difficult to solve fine-grained combinatorial optimization problems, the finer the granularity is, the higher the solution complexity is. For an intelligent evolutionary interaction environment with granularity value $g$ and the number of environmental factors $n$, the total number $k$ of all allocation schemes is

$$k = \boldsymbol{C}_{g+n-1}^{n-1} = \frac{(n-1)!}{(g+n-1)!\, g!} \tag{1}$$

The set of all interaction schemes is

$$T = \{T_1, T_2, \ldots, T_k\} \tag{2}$$

where each element in $T$ represents an allocation strategy,

$$T_i = \{a_1^i, a_2^i, \ldots, a_n^i\},\ \sum_{j=1}^{n} a_j^i = g,\ a_j^i \in \mathbf{N} \tag{3}$$

If $t$ is controlled within a small range, then the global optimal allocation strategy $\hat{T}_i$ can be sought through the traversal. If $t$ is too large, then some combinatorial optimization algorithm should be sought to find the optimal solution in a short time. In either way, $\hat{T}_i$ must satisfy

$$\operatorname{argmin}\ f(A, E, \hat{T}_i) \tag{4}$$

where $A$ represents the state of the agent and contains $m$ item attributes. $E$ represents the state of the interactive environment, which contains $n$ environment factors. The attributes contained in each environmental factor can be converted to the state attributes of the agent. $f(\cdot)$ represents a measure of intelligence, which is represented as

$$f(A, E, \hat{T}_i) = \mathrm{SD}(A + h(E, \hat{T}_i), \bar{A}_s) \tag{5}$$

where $\mathrm{SD}(\cdot)$ stands for the standard eleviation, $\bar{A}_s$ represents the standard state of the agent, and $h(\cdot)$ represents a conversion scheme. Through the conversion $h(\cdot)$, the attributes of the environmental factor can be directly calculated with the state $A$. As time goes on, the state of the agent at time $t$ is $A_t$, the evolution state of the agent at the next moment is

$$A_{t+1} = A_t + h(E, \hat{T}_i^t) - \bar{A}_s \tag{6}$$

### 2.2 Agent and interactive environment settings

For real-world scenarios, the formulas presented in the previous subsection need to be instantiated. In this study, the health condition of an agent is regarded as its intelligence according to the diet collocation scene. $\bar{A}_s$ represents a standard pursued by an agent, which is set in the instance as

$$\bar{A}_s = [82.6, 2100, 55, 300, 1000] \tag{7}$$

These scalars represent standards that an agent strives to meet after each evolution, and that each agent needs to deduct before it evolves. These scalars represent the number of units of protein, calories, fat, carbohydrates, and sodium that an agent needs to ingest for each evolution. For the state of the interactive environment, $E$ is defined as a combination of $n = 6$ kinds of foods,

$$E = [E_1, E_2, E_3, E_4, E_5, E_6] \tag{8}$$

To simplify the calculation, we define the element $E_i$ in the environment variable $E$ to have only 1 and 0 values. For example, when $E_3 = 1$ represents the positive foods and their indexes which ID = 3, and when $E_3 = 0$ represents the negative foods and their indexes, ID = 3. Tables 1 and 2 show the details of six positive foods and six negative foods, respectively. All these data were collected from the Internet.

In this case, $g = 20$ and the agent's total appetite $p = 2000$, $h(\cdot)$ is expressed as

$$h(E, \hat{T}_i) = \frac{\hat{T}_i E p}{g} \tag{9}$$

Here, the 2000 units of food are divided into 20 equal portions, and the 20 portions are divided into 6 environmental factors according to $\hat{T}_i$. In addition, based on the above setting, $k = 53\,130$, we can find the globally optimal $\hat{T}_i$ using the traversal.

Finally, the agent will use greedy strategies to ingest every food resource in the environment under the quantitative conditions allowed by the computer. The whole process adopts the traversal method.

## 3 Evolutionary Experiments and Observations

Based on the evolution principle and parameter setting discussed in Section 2, this section presents the evolution and observation of the evolutionary behavior.

First, to verify the reliability of our method, we set $\bar{A}_s'$ as the fluctuating state. At different evolution stages, $\bar{A}_s'$ fluctuates within a specific range. Figure 2 shows four fluctuating conditions of different degrees. Among them, $\bar{A}_s'$ of each iteration changes and follows

**Table 1 Positive foods and their nutrition indices. Different nutritional indexes in this table have different measurements. This paper only refers to numerical values.**

| ID | Positive food | Protein | Calorie | Fat | Carbohydrate | Sodium |
|---|---|---|---|---|---|---|
| 1 | Beef | 0.1573 | 3.32 | 0.233 | 0 | 0.66 |
| 2 | Fish | 0.1647 | 0.99 | 0.130 | 0.001 | 1.03 |
| 3 | Egg | 0.1284 | 1.39 | 0.086 | 0.024 | 1.74 |
| 4 | Mushroom | 0.0270 | 0.24 | 0.001 | 0.041 | 0.20 |
| 5 | Grain | 0.0213 | 1.16 | 0.003 | 0.259 | 0.22 |
| 6 | Vegetable | 0.0165 | 0.23 | 0.003 | 0.038 | 0.56 |

**Table 2 Negative foods and their nutrition indices. Different nutritional indexes in this table have different measurements. This paper only refers to numerical values.**

| ID | Negative food | Protein | Calorie | Fat | Carbohydrate | Sodium |
|---|---|---|---|---|---|---|
| 1 | Chip | 0.075 | 5.48 | 0.376 | 0.530 | 4.000 |
| 2 | Chocolate | 0.043 | 5.89 | 0.401 | 0.534 | 1.118 |
| 3 | Cake | 0.072 | 3.79 | 0.139 | 0.565 | 0.807 |
| 4 | Ice cream | 0.035 | 2.20 | 0.141 | 0.200 | 0.542 |
| 5 | Popcorn | 0.108 | 4.06 | 0.134 | 0.704 | 0 |
| 6 | Instant noodle | 0.095 | 4.73 | 0.211 | 0.616 | 14.44 |

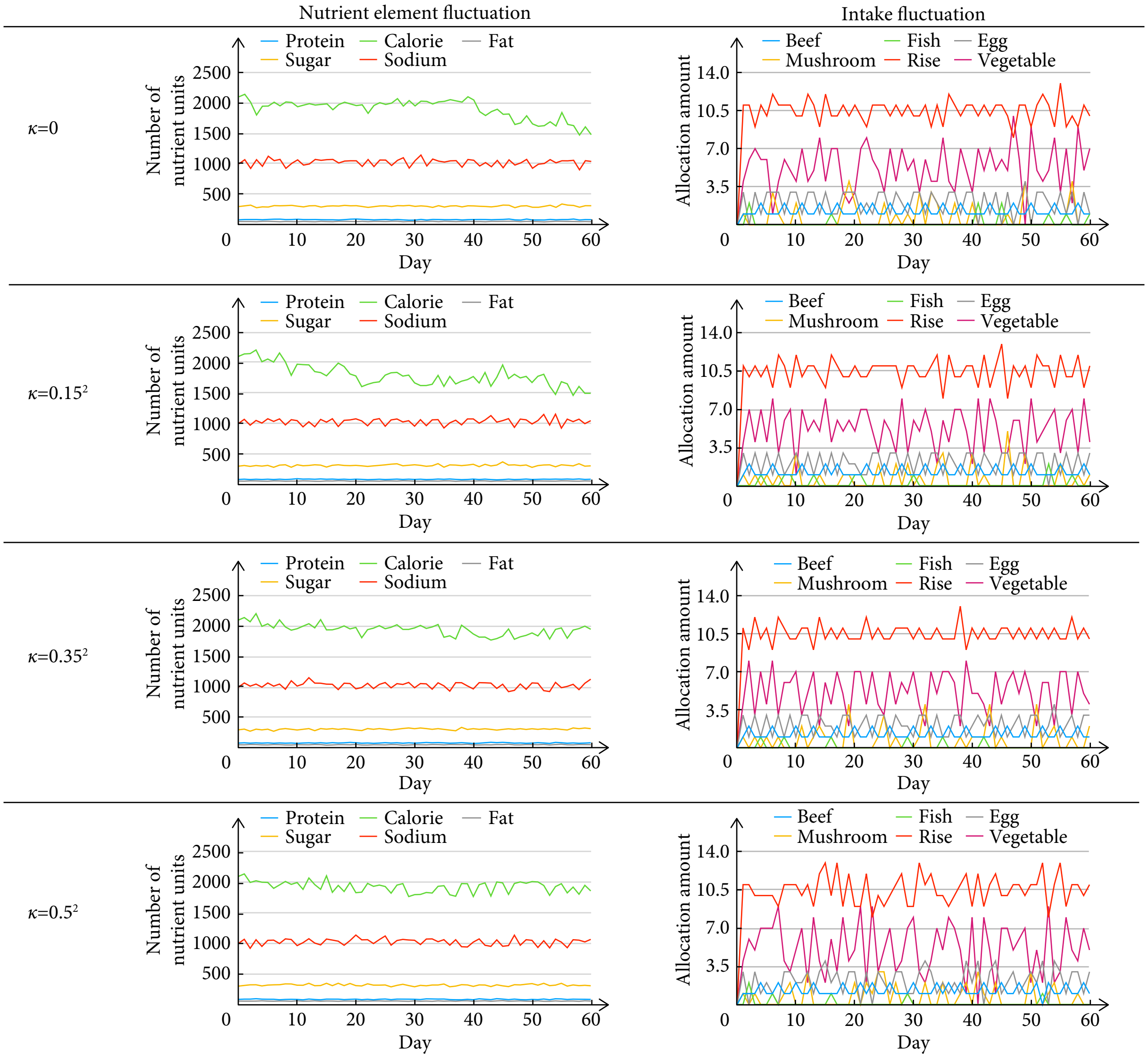


**Fig. 2 Evolution process under different fluctuation coefficients $\kappa$. A great fluctuation coefficient does not have a noticeable effect on the evolution.**

$$\bar{A}_s' = \bar{A}_s + U(-\kappa, \kappa) \tag{10}$$

where $U(a, b)$ represents the uniform distribution between $a$ and $b$, and $\kappa$ is the fluctuation coefficient of fluctuation. Figure 2 shows that all elements of environmental factor $E$ are 1 and evolve 60 times in a row. The fluctuation curve shown in the left column is the changes in the five agent attributes in the evolution process, whereas the fluctuation curve shown in the right column is the balance adjustment changes in $\hat{T}_i$ in the evolution process. The results show that our proposed evolutionary framework can support the fluctuation of the evolutionary goals at each stage within the $\pm 25\%$ range. In addition, it can be seen from Fig. 2 that when the volatilization coefficient $\kappa = 0.5^2$, $T^s$ fluctuates greatly in the controlled experiment. Hence, the proposed method can adapt to a large range of volatility range.

Then, we examined the evolution of several different sets of interaction factors. In the evolution shown in Fig. 3, the leftmost number represents the code for the interaction factor $E$ (e.g., 101010 represents the value 1 of $E_1$, 0 of $E_2$, 1 of $E_3$, and so on). The waveform on the left shows the changes in the agent's state $A$ at different times. The waveform on the right shows the change in $\hat{T}_i$. The whole evolution happens at $\kappa = 0$. In particular, the evolution environment under some interaction factors is too harsh. When the sum of the standard deviation between the five attributes of agent and $\bar{A}_s$ is greater than 99, we will stop evolution. Based on the observation of different environmental factors $E$ in Fig. 3, not all environmental factors are evolution friendly. Because even if we get the best solution at every evolution stage, we still cannot perform as well in all evolutionary contexts. However, in Fig. 2, the interaction frequency of $E_2 = 1$ is extremely low in the evolution process, with only 0.13 unit of interaction per evolution on average. Therefore, even if $E_2 = 1$ is a positive food, it has been coupled with other food combinations, so replacing it may be a good choice. Even if replaced with a certain negative food, it will not have a huge impact on the evolution course.

Figure 4 shows the change in the sum of the standard deviations of the agent state and standard state $A_s$ in each evolutionary environment. The sum of the standard deviations of the environments 101010 and 000000 exceeded 99 in less than 60 evolutions. The sum of the standard deviations of environments 100011 and 111011 also presents an upward trend. The two evolutionary environments, 111111 and 101111, are suitable for evolution because they do not show a significant upward trend in Fig. 4.

The agent's total appetite $p$ is also an evolutionary factor that

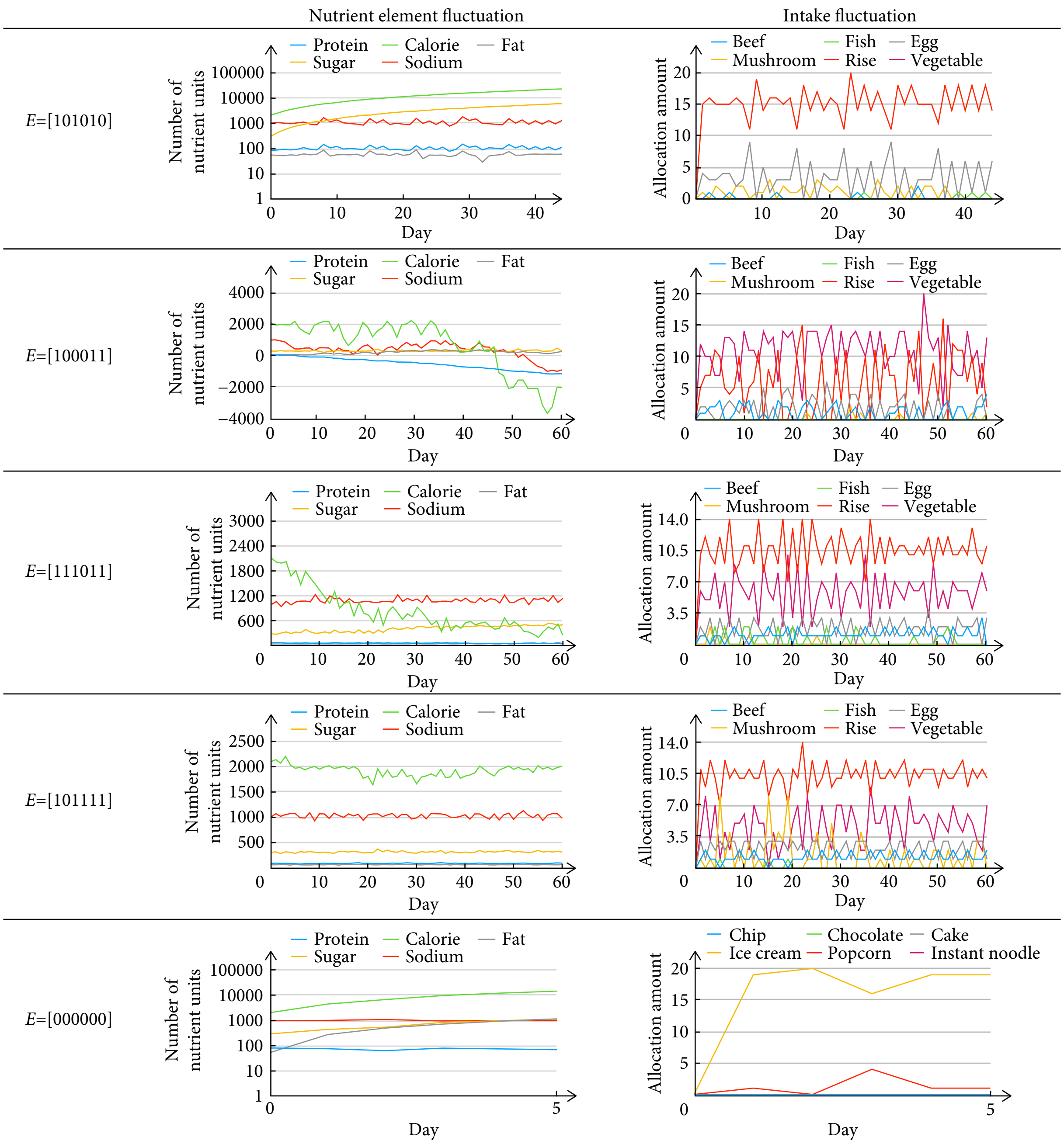


**Fig. 3 State change and evolutionary strategy selection of agents in different evolutionary environments.**

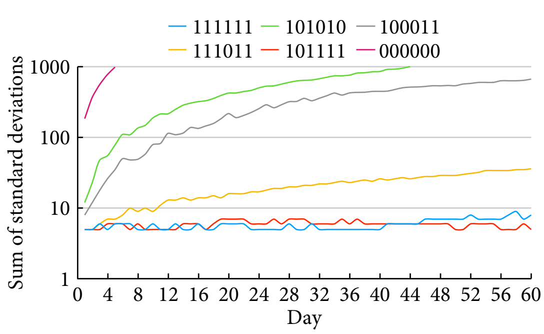


**Fig. 4 Comprehensive state changes of agents in different evolutionary environments.**

cannot be ignored. The evolutionary examples in Figs. 2–4 are all performed with $p$ = 2000. We believe that changing the value of $p$ can alleviate the disadvantage of the evolutionary environment. In our experiment, the range of $p$ is expanded to (1000, 3000), and the six intelligent evolutionary environments in Fig. 4 are evolved 30 times with a step length of 50. Then, the results are recorded. Figure 5 shows the recorded results. For the three intelligent evolutionary environments 101010, 100011, and 111011, a proper adjustment of the value of $p$ can promote evolution, but it can only maintain this state in a small range. The evolutionary environments 111111 and 101111 maintain a good evolutionary state within a relatively large appetite range. This finding indicates that the two environments are most conducive to agent evolution.

## 4 Conclusions and Future Works

In this study, a balanced diet is taken as an example, and its relevant data are used as support to make agents exhibit independent evolutionary behaviors. In this mode, the agent generates "actions" to different environmental factors around it,

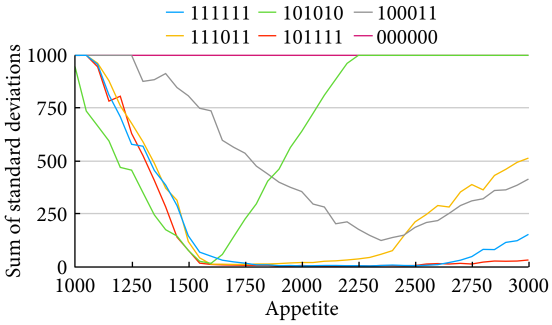


**Fig. 5 Relationship between intelligence and appetite in different evolutionary environments. This analysis can be used to determine whether an agent is suitable for intelligent evolution from another point of view.**

and the corresponding environmental factors feedback some rewards to the agent. Agents can find the best way to interact with various environmental factors through the traversal. Hence, the environment of the agent can reflect its intelligence quantity, which is reflected by the changing trend of the intelligence quantity in a certain period. The intelligence of the agents shows different evolutionary trends in the different eating and drinking scenarios we set up. At the same time, the proposed method proves whether the interactive environment is conducive to the evolution of agents. In the future, we will study the evolution of intelligence in more heterogeneous scenarios.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 62072440) and the Beijing Natural Science Foundation (No. 4202072).

## Dates

Received: 3 March 2021; Revised: 27 February 2022; Accepted: 28 February 2022

## References

[1] D. L. Dowe and J. Hernández-Orallo, IQ tests are not for machines, yet, *Intelligence*, vol. 40, no. 2, pp. 77–81, 2012.

[2] A. M. Turing, Computing machinery and intelligence, *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[3] G. Longo, Laplace, turing and the "imitation game" impossible geometry, in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, R. Epstein, G. Roberts, and G. Beber, eds. Dordrecht, the Netherlands: Springer, 2009, pp. 377–411.

[4] J. Hernández-Orallo and D. L. Dowe, Measuring universal intelligence: Towards an anytime intelligence test, *Artif. Intell.*, vol. 174, no. 18, pp. 1508–1539, 2010.

[5] M. Dabaghchian, S. S. Liu, A. Alipour-Fanid, K. Zeng, X. H. Li, and Y. Chen, Intelligence measure of cognitive radios with learning capabilities, in *Proc. 2016 IEEE Global Communications Conf.* (*GLOBECOM*), Washington, DC, USA, 2016, pp. 1–6.

[6] W. Ji, J. Liu, Z. W. Pan, J. C. Xu, B. Liang, and Y. Q. Chen, Quality-time-complexity universal intelligence measurement, *Int. J. Crowd Sci.*, vol. 2, no. 1, pp. 18–26, 2018.

[7] S. S. Xiao, C. P. Wei, and M. Dong, Crowd intelligence: Analyzing online product reviews for preference measurement, *Inf. Manage.*, vol. 53, no. 2, pp. 169–182, 2016.

[8] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone, Evidence for a collective intelligence factor in the performance of human groups, *Science*, vol. 330, no. 6004, pp. 686–688, 2010.

[9] S. Costantini, Defining and maintaining agent's experience in logical agents, in *Proc. 7th Latin American Workshop on Non-Monotonic Reasoning*, Toluca, Mexico, 2011, pp. 151–165.

[10] Y. H. Chang, T. Y. Lu, and R. J. Fang, An adaptive e-learning system based on intelligent agents, in *Proc. 6th Conf. on WSEAS Int. Conf. on Applied Computer Science*, Hangzhou, China, 2007, pp. 200–205.

[11] J. Duncan, R. J. Seitz, J. Kolodny, D. Bor, H. Herzog, A. Ahmed, F. N. Newell, and H. Emslie, A neural basis for general intelligence, *Science*, vol. 289, no. 5478, pp. 457–460, 2000.

[12] D. F. Hougen and S. N. H. Shah, The evolution of reinforcement learning, in *Proc. 2019 IEEE Symp. Series on Computational Intelligence* (*SSCI*), Xiamen, China, 2020, pp. 1457–1464.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[14] R. Lowe and T. Ziemke, Exploring the relationship of reward and punishment in reinforcement learning, in *Proc. 2013 IEEE Symp. on Adaptive Dynamic Programming and Reinforcement Learning* (*ADPRL*), Singapore, 2013, pp. 140–147.