Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

*Perspective:*

# Existence and practice of gaming: thoughts on the development of multi-agent system gaming*

Qi DONG[†1], Zhenyu WU[†1,2], Jun LU[†‡1], Fengsong SUN[1,3],
Jinyu WANG[1,3], Yanyu YANG[1], Xiaozhou SHANG[1]

[1]*China Academy of Electronics and Information Technology, Beijing 100049, China*
[2]*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*
[3]*School of Information and Communication Engineering,*
*Beijing University of Posts and Telecommunications, Beijing 100876, China*
[†]E-mail: dongqiouc@126.com; zhenyuwu1993@126.com; lujun_caeit@126.com

Received Dec. 30, 2021; Revision accepted May 5, 2022; Crosschecked June 1, 2022

https://doi.org/10.1631/FITEE.2100593

Game is a universal being in the universe. Starting with human understanding of the game process, we discuss the existence and practice of gaming, expound challenges in multi-agent gaming, and put forward a theoretical framework for a multi-agent evolutionary game based on the idea of evolution and system theory. Taking the next-generation early warning and detection system as an example, we introduce the applications of multi-agent evolutionary game. We construct a multi-agent self-organizing game decision-making model and develop a multi-agent method based on reinforcement learning, which are significant in studying organized and systematic game behaviors in a high-dimensional complex environment.

## 1 Introduction

Gaming is everywhere. From biological population to human society, from tribal conflict to superpower games, and from the exchange of goods to financial trade, all these scenarios are permeated with the idea of game. Game has become a universal being in the universe (Heidegger, 2013). There has been an in-depth development of human understanding of game. The early research on game theory summarized mainly the experience of war, chess, and card activities (Kant, 2020). *Sun Tzu's Art of War* is one of the earliest works on game theory. Von Neumann (1928) proved the mini-max theorem of the zero-sum game and established a mathematical research framework of game theory, marking the birth of classical game theory. In 1944, the emergence of the epoch-making masterpiece *Theory of Games and Economic Behavior* laid a theoretical foundation for applying classical game theory in economics, and represented a breakthrough in the development of basic assumptions and analysis paradigms of traditional economics (Von Neumann and Morgenstern, 2007). The Nash equilibrium theory (Nash, 1950) made game theory a widely used analysis tool.

After 1970, the ideas of incomplete information games and bounded rational decision-making were integrated, and the practical research and application scope of game theory was dramatically expanded. With the development of computer technology, using trial-and-error data to establish decision-making methods has become a new idea to solve

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (No. 61803353)
ORCID: Qi DONG, https://orcid.org/0000-0002-1215-7588

gaming problems. For example, AlphaGo, based on intensive learning and training, defeated the world champion Lee Se-Dol in the game of Go (Silver et al., 2016). In recent years, game theory has been continuously improved and has gradually become an analytical framework for solving problems in many fields. For example, political struggle, military confrontation, economic analysis of market behavior (Abu Turab Rizvi, 2007), exploration of cooperative mechanisms in biological populations (Archetti and Pienta, 2019), and policy-making in social governance among significant powers (Wang et al., 2021) can be analyzed or conducted with the game theory. These games have complex institutionalized and systematized characteristics.

Institutionalized and systematized gaming research has expanded from individuals to groups. When groups reach a particular scale, they can often exhibit characteristics that are different from those of individuals (Cavagna et al., 2010; Alsheikh et al., 2015; Hayat et al., 2016). How to analyze and use these characteristics has been of great concern, and the concept of multi-agent system (MAS) is prompted, which is defined as a complex system composed of multiple agents that interact with environments (Shoham and Leyton-Brown, 2008). MAS gaming (MASG) provides a theoretical framework for studying the above issues.

Although MASG has been significantly improved in the past decades, it still faces many challenges:

(1) The environment, in which the system is located, is complex and changeable. It is challenging to model the environment directly and predict the environment's response to the agents' actions accurately.

(2) The system is heterogeneous, and heterogeneous individuals have different decision spaces; it is far from easy for complex systems to achieve coordinated control.

(3) Due to the limitations of distance, power, and other factors, the perception of agents is limited. So, the generated situation is incomplete and inconsistent.

(4) The computing power of a single agent in a system is limited. So, it is difficult to manage a large amount of data generated by the system in the gaming process and to generate the best decision in real time.

Evolutionary game theory (EGT) (Smith, 1982), inspired by Darwin's theory of evolution, provides efficient approaches for complex problems in situations of incomplete information and bounded rationality. Evolutionary thought has introduced a new idea for solving the above issues. Multi-agent gaming (MAG) algorithms based on evolutionary thought have become a hotspot in gaming research (Nowak, 2006; Hilbe et al., 2018; Omidshafiei et al., 2019; Gupta et al., 2021). New algorithms, such as multi-agent reinforcement learning (MARL) (Shao et al., 2019), the ant colony optimization (ACO) algorithm, and the particle swarm optimization (PSO) algorithm (Liu ZA and Nishi, 2022), have achieved significant success and have been applied in many fields. For example, these algorithms study cooperation in society (Nowak, 2006; Hilbe et al., 2018), cooperative decision-making in multi-party games (Shao et al., 2019), cooperative control schemes in intelligent transportation (Li et al., 2019), and self-organizing game decision-making in distributed early warning and detection.

Although these algorithms have achieved significant results in many applications, the theoretical framework is unclear. They have only the correctness of formal logic but no truth of dialectical reasoning. They are like water without a source or a tree without a root, and cannot reveal the truth behind multi-agent evolutionary gaming (MAEG) (Heidegger and Mörchen, 1988). Therefore, using the ternary system theory, this paper attempts to put forward the theoretical research framework of MAEG and apply it to early warning and detection. By exploring the essence of MAEG in practice, we hope to trigger more scholars' thinking and research and promote the development of this field.

## 2 Theoretical framework of multi-agent evolutionary gaming

In this section, we propose a theoretical framework for MAEG, as shown in Fig. 1. According to the ternary system theory, system $\mathcal{S}$ is an organic whole composed of elements $\mathcal{E}$, relations $\mathcal{R}$, and laws $\mathcal{L}$ (Xu, 2000; Lu and Shan, 2020). We assume that elements represent the collection of agents $N_t$ (which denotes the agent set, including homogeneous and heterogeneous agents) and environment $E_t$ in an MAS. The symbol $t$ indicates that agents and the environment continuously evolve with time. Relations $\mathcal{R}$

in an MAS represent the collection of the interaction relationships between not only agents, but also the system and environment. Laws $\mathcal{L}$ in an MAS are goal-oriented and include the internal and external power.

Generally speaking, agents' strategies $X_t = \{X_t^1, X_t^2, \cdots, X_t^N\}$ and agents' relationships in an MAS will constantly evolve under the joint promotion of internal power $D_t^{\text{in}}$ and external power $D_t^{\text{out}}$. $D_t^{\text{out}}$ denotes the change rules of the environment, the evolution of the environment, and the interaction between the environment and the system. $D_t^{\text{in}}$ denotes the change rules of agents based on the internal relationships of the system, including the evolutionary mechanism of agents in the game process.

Under the constraint and drive of evolutionary game laws $\mathcal{L}$, MAEG is directional in obtaining the overall income $C_t$ of MAS:

$$C_t = \mathcal{L} < D_t^{\text{in}}, D_t^{\text{out}}, N_t, E_t, X_t > . \tag{1}$$

Through MAEG, the system can evolve from disorder to order, and the steady-state value of income $C_k$ can be obtained as follows:

$$||C_k - C_{k-1}|| \leq \epsilon,\ \epsilon > 0. \tag{2}$$

## 3 Practical research on MAEG

Early warning and detection is significant in rescue and relief work, urban public security, and other situations. Nowadays, a high-dimensional and complex environment raises higher requirements in early warning and detection systems. In this section, we take early warning and detection as an example and focus on the theoretical exploration and practical application of MAEG.

The next-generation early warning and detection system will be distributed, unmanned, and intelligent (Liu M and Lu, 2015). It is composed of multiple, cooperative, distributed nodes with independent

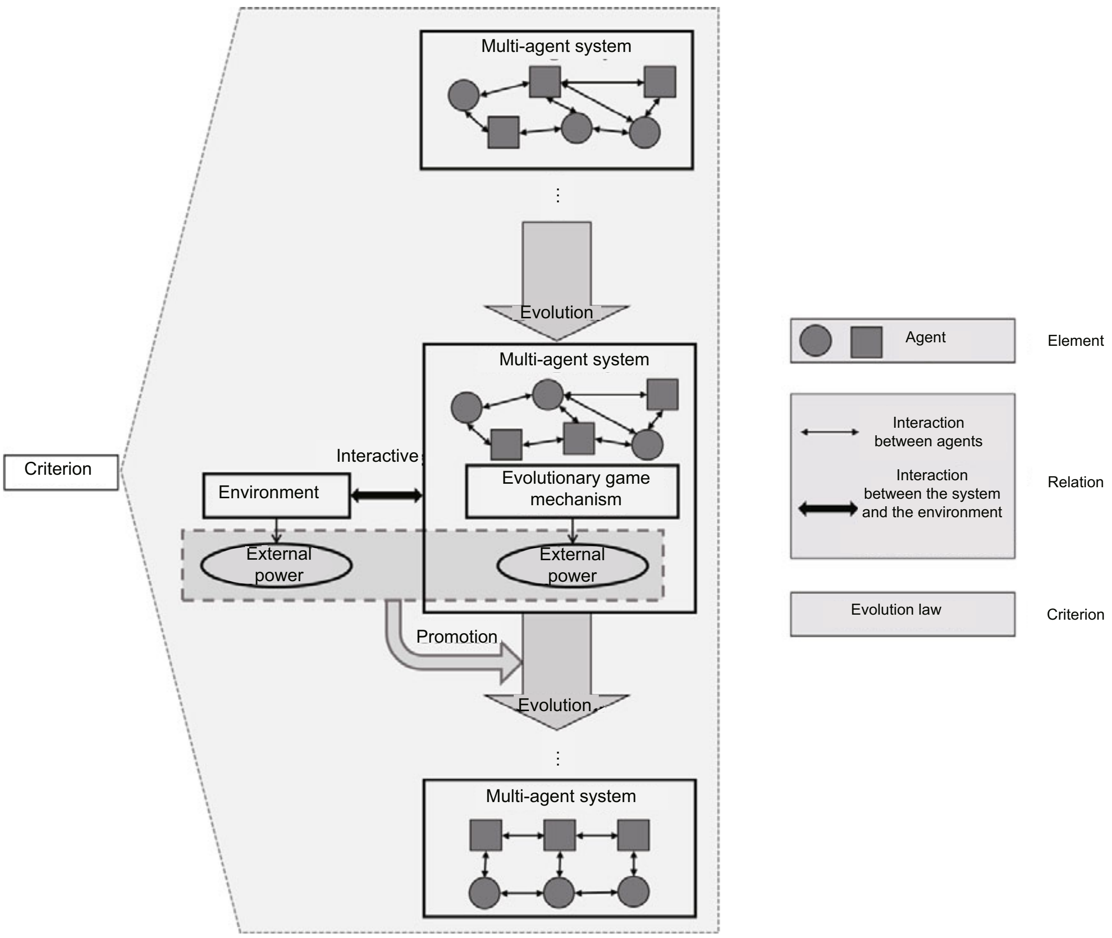


**Fig. 1 Theoretical framework of multi-agent evolutionary gaming (MAEG)**

sensing, decision-making, and action. The system achieves the cooperative sensing and recognition of the environment through the coordinated control and information fusion of nodes. Each node must adapt to the unknown and changing external environment in a distributed early warning and detection system. At the same time, the system must allocate cooperative detection resources among nodes, including perception, communication, and computing. Therefore, a distributed early warning and detection system is a typical MAEG system. Based on the MAEG theory, we study multi-agent collaborative detection methods, construct a multi-agent self-organizing game decision-making model, and establish "internal power" of system evolution using data-driven decision learning methods. In this way, a next-generation early warning and detection system is formed that can adapt to environmental changes, has highly reliable operations, and can create an accurate and effective unified situation, as shown in Fig. 2.

### 3.1 Multi-agent self-organizing game decision-making model

We construct a self-organizing game decision-making model in a distributed early warning and detection system using the MAEG theoretical framework, as shown in Fig. 3. Considering a system composed of $N$ agents, the self-organizing game decision-making model is formalized by the tuple $(\mathcal{N}, S, \mathcal{G}, A, F, \mathit{\Pi})$, where $\mathcal{N} = \{1, 2, \cdots, N\}$ denotes the set of agents, $S$ is the system's situation information, and $S_t = \{s_t^1, s_t^2, \cdots, s_t^N\}$ is the current situation information of agents. $\mathcal{G}$ represents the time-varying communication topology of multiple agents in the system in the self-organizing network architecture. $A = \prod_{i=1}^{N} A_i$ is the product of finite action spaces of all agents, known as the joint action space, and $A_i$ denotes the action set of agent $i$. $F = \{F_1, F_2, \cdots, F_N\}$ is the set of agents' fitness in the system, where $F_i$ represents the adaptability of agent $i$ to the task requirements. $\mathit{\Pi} = \{\mathit{\Pi}_1, \mathit{\Pi}_2, \cdots, \mathit{\Pi}_N\}$ denotes the policy set of all agents and $\mathit{\Pi}_i = \{\pi_{i1}, \pi_{i2}, \cdots, \pi_{ik}\}$ indicates the optional policy set of agent $i$. The agents interact with the environment and update policies according to the following protocol: At time step $t$ ($t \in \mathbb{N}$), agent $i$ predicts situation information $s_{t+1}^i$ drawn from the current situation information $s_t^i$. $s_t^i$ and $s_{t+1}^i$ then work on $F_i(\cdot)$ to give an evolution of the current policy $\pi_{ij}$ ($i = 1, 2, \cdots, N, j = 1, 2, \cdots, k$). Considering the environmental information $o_i$, agent $i$ takes action $a_t^i \in A_i$ drawn from its joint policy $\pi_{ij}$, influenced by the environmental information, current

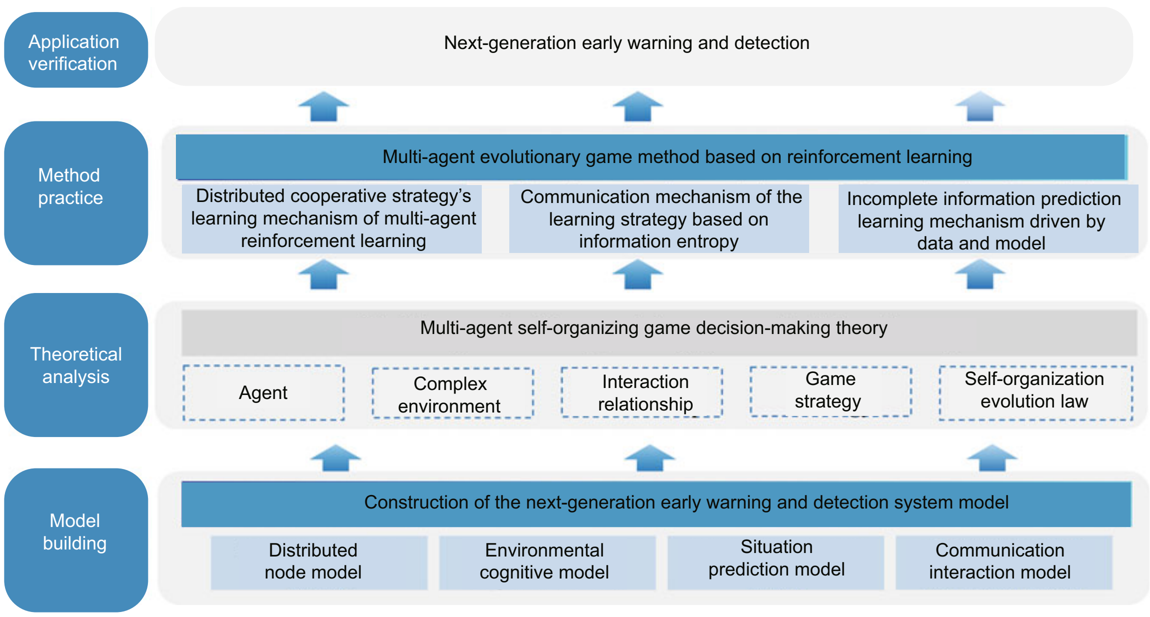


**Fig. 2 Applications of MAEG in early warning and detection**

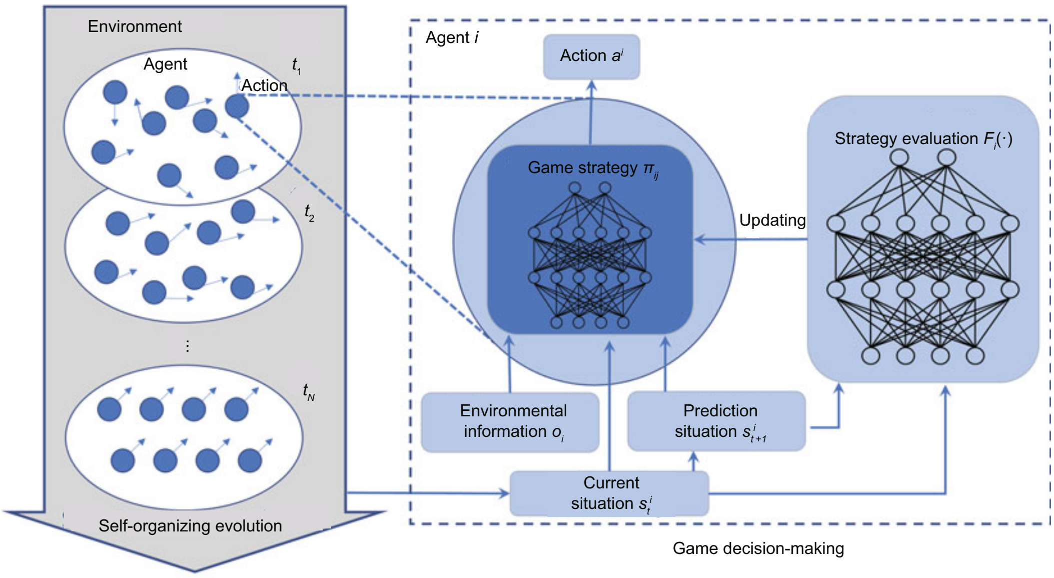


**Fig. 3 Self-organizing game decision-making model**

and predicted situation information $s_t^i$, $s_{t+1}^i$, and fitness function $F_i(\cdot)$. After multiple rounds of environmental interaction and sample accumulation, the agents perform policy updating and optimization. In the early warning and detection system, the internal power is composed of the current situation $s_t^i$, the next situation $s_{t+1}^i$, and agents' fitness $F_i(\cdot)$. At the same time, the environmental cognitive results $o_i$'s constitute the external power of the system. The internal power and external power are aggregated to guide the agent iteration for policy evaluation and updating, drive the individuals to complete the self-organizing evolution of group actions based on the game strategy, and finally form the ability of the nodes to adapt to the environment and efficiently cooperate in the distributed early warning and detection system.

### 3.2 MAEG method based on reinforcement learning

Based on the self-organizing game decision-making model, we further propose an MAEG method based on reinforcement learning, as shown in Fig. 4, which is used to solve the problems of multi-agent cooperative decision-making and effective real-time communication in a complex dynamic environment in the next-generation early warning and detection system. Considering the distributed cooperative strategy learning mechanism of MARL, we first construct the temporal difference optimization objective $\mathcal{L}_{\mathrm{TD}}(\boldsymbol{\theta})$, based on the joint state value function $Q^{\mathrm{tot}}$ and agent state value function $Q^i$ as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{TD}}(\boldsymbol{\theta}) = \Big[ & r + \gamma \max_{a_{t+1}} Q^{\mathrm{tot}}(a_{t+1}, s_{t+1}; \boldsymbol{\theta}^-) \\ & - Q^i(a_t, s_t; \boldsymbol{\theta}^-)\Big]^2,\end{aligned} \tag{3}$$

where $r$ is the reward function, $\gamma \in (0,1)$ is the discount factor, and $\boldsymbol{\theta}$ is the parameter vector of agents' policy $\mathit{\Pi}$. At time step $t$ ($t \in \mathbb{N}$), agent $i$ executes $\varepsilon$ strategy based on $Q^i$ and obtains action $a^i$, denoted as

$$a^i = \begin{cases} \arg\max\limits_{a \in A_i} Q^i, & p < \varepsilon, \\ \forall a \in A_i, & p \geq \varepsilon, \end{cases} \tag{4}$$

which is used to improve the efficiency of multi-agent self-organizing games. This optimization objective solves the collaborative planning, collaborative control, and real-time decision-making problems of large-scale agents. Furthermore, we establish the communication mechanism of the learning strategy based on information entropy. The optimization goal of minimum information entropy $\mathcal{L}_{\mathrm{G}}(\boldsymbol{\theta}_{\mathrm{G}})$ is constructed by establishing an intelligent communication model:

$$\mathcal{L}_{\mathrm{G}}(\boldsymbol{\theta}_{\mathrm{G}}) = E_{\tau \sim D} \sum_{i \neq j} \left[ I_{\boldsymbol{\theta}_{\mathrm{G}}}(a^j; m_{ij}) - \beta H_{\boldsymbol{\theta}_{\mathrm{G}}}(m_{ij}) \right], \tag{5}$$

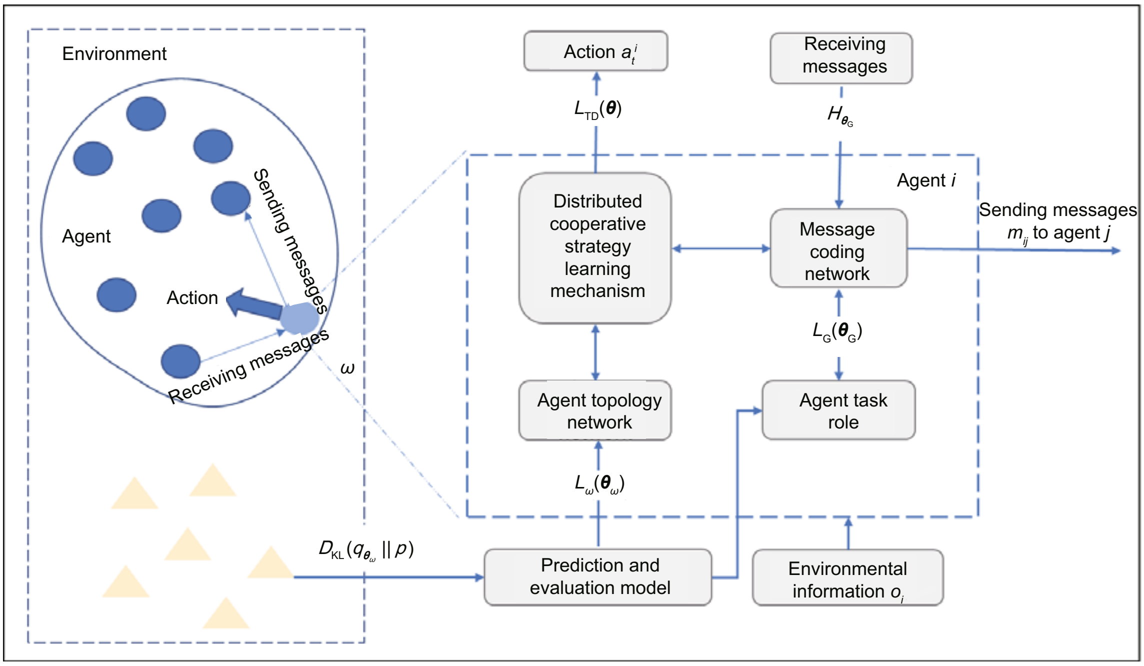


**Fig. 4 A multi-agent evolutionary game method based on reinforcement learning**

where $I_{\boldsymbol{\theta}_{\rm G}}$ refers to the mutual information, $m_{ij}$ is the message sent by agent $i$ to agent $j$, $H_{\boldsymbol{\theta}_{\rm G}}$ is the information entropy, and $\boldsymbol{\theta}_{\rm G}$ encodes network parameters for messages. The subscript "G" refers to the topological network of multiple agents. $\mathcal{L}_{\rm G}(\boldsymbol{\theta}_{\rm G})$ is used to maximize the interactive information between messages and action selection to realize full expression of messages. In addition, aiming at the problem of unknown environmental information in complex dynamic environments, we use an incomplete information prediction learning mechanism driven by data and model. This allows us to establish a predictive evaluation model and construct a prediction learning objective driven by data and model. The prediction learning objective $\mathcal{L}_\omega(\boldsymbol{\theta}_\omega)$ takes the following form:

$$\mathcal{L}_\omega(\boldsymbol{\theta}_\omega) = D_{\rm KL}\Big(q_{\boldsymbol{\theta}_\omega}\left(s_{t+1}|s_t, a_t\right) || p\left(s_{t+1}|s_t, a_t\right)\Big), \tag{6}$$

where $\omega$ is the prediction parameter, $D_{\rm KL}$ is the Kullback–Leibler (KL) divergence, $q_{\boldsymbol{\theta}_\omega}(\cdot)$ is the environmental model prediction network, and $p(\cdot)$ refers to the dynamic environmental model. According to the mechanisms mentioned, the system can realize environment modeling and evolutionary learning with incomplete information, and infer agent topology network task roles. So, the agents can grasp as much comprehensive information as possible and realize more effective decision-making. After constant policy updating and optimization, the early warning and detection system can finally achieve self-organizing evolution and achieve a systematic capability that is different from that of a single node.

## 4 Conclusions and prospects

Taking the widespread existence of games in the universe as the starting point, in this paper we describe the process of human's understanding of gaming, expound challenging problems of the multi-agent game process, put forward the theoretical MAEG research framework, and introduce MAEG's practical applications using the next-generation early warning and detection system as an example.

MAEG theory is significant for studying institutionalized and systematized gaming in high-dimensional complex environments. Existing research has made a preliminary exploration in this field. However, recent research focuses on evolving games with faster convergence, more stability, and better performance based on definable criteria to ensure formal logic correctness. The criteria, based on which concepts are formed and used for judgment and reasoning decisions, are difficult to describe for systematic and organizational games. Existing methods have difficulty in solving such problems, and the truth of dialectical logic needs to be studied to further explore game criteria. Human inquiry about thinking has never stopped, and noetic science is currently the key to tackling challenges. Although studying thinking is difficult, we feel that noetic science is the direction of such research and the ultimate goal of games.

Our view aims at starting thinking and discussing the multi-agent evolutionary game from different dimensions.

## Contributors

Qi DONG and Jun LU designed the research. Zhenyu WU and Fengsong SUN drafted the paper. Yanyu YANG and Xiaozhou SHANG helped organize the paper. Jinyu WANG revised and finalized the paper.

## Compliance with ethics guidelines

Qi DONG, Zhenyu WU, Jun LU, Fengsong SUN, Jinyu WANG, Yanyu YANG, and Xiaozhou SHANG declare that they have no conflict of interest.

## References


Abu Turab Rizvi S, 2007. Aumann's and Schelling's game theory: the Nobel Prize in Economic Science, 2005. *Rev Polit Econ*, 19(3):297-316. https://doi.org/10.1080/09538250701452990

Alsheikh MA, Hoang DT, Niyato D, et al., 2015. Markov decision processes with applications in wireless sensor networks: a survey. *IEEE Commun Surv Tutor*, 17(3):1239-1267. https://doi.org/10.1109/COMST.2015.2420686

Archetti M, Pienta KJ, 2019. Cooperation among cancer cells: applying game theory to cancer. *Nat Rev Cancer*, 19(2):110-117. https://doi.org/10.1038/s41568-018-0083-7

Cavagna A, Cimarelli A, Giardina I, et al., 2010. Scale-free correlations in starling flocks. *Proc Nat Acad Sci USA*, 107(26):11865-11870. https://doi.org/10.1073/pnas.1005766107

Gupta A, Savarese S, Ganguli S, et al., 2021. Embodied intelligence via learning and evolution. *Nat Commun*, 12(1):5721. https://doi.org/10.1038/s41467-021-25874-z

Hayat S, Yanmaz E, Muzaffar R, 2016. Survey on unmanned aerial vehicle networks for civil applications: a communications viewpoint. *IEEE Commun Surv Tutor*, 18(4):2624-2661. https://doi.org/10.1109/COMST.2016.2560343

Heidegger M, 2013. Sein und Zeit. De Gruyter, Germany (in German).

Heidegger M, Mörchen H, 1988. Vom Wesen der Wahrheit Zu Platons Höhlengleichnis und Theätet. V. Klostermann, Frankfurt, USA (in German).

Hilbe C, Šimsa Š, Chatterjee K, et al., 2018. Evolution of cooperation in stochastic games. *Nature*, 559(7713):246-249. https://doi.org/10.1038/s41586-018-0277-x

Kant I, 2020. Kritik der Reinen Vernunft. BoD–Books on Demand, Germany (in German).

Li L, Wang X, Wang KF, et al., 2019. Parallel testing of vehicle intelligence via virtual-real interaction. *Sci Robot*, 4(28):eaaw4106. https://doi.org/10.1126/scirobotics.aaw4106

Liu M, Lu J, 2015. Overview of AEW character and development trend. *J China Acad Electron Inform Technol*, 10(3):278-282 (in Chinese).

Liu ZA, Nishi T, 2022. Strategy dynamics particle swarm optimizer. *Inform Sci*, 582:665-703. https://doi.org/10.1016/j.ins.2021.10.028

Lu J, Shan BN, 2020. Information system development thinking. *J Electr Electron Edu*, 42:1 (in Chinese).

Nash JFJr, 1950. Equilibrium points in $n$-person games. *Proc Nat Acad Sci USA*, 36(1):48-49. https://doi.org/10.1073/pnas.36.1.48

Nowak MA, 2006. Five rules for the evolution of cooperation. *Science*, 314(5805):1560-1563. https://doi.org/10.1126/science.1133755

Omidshafiei S, Papadimitriou C, Piliouras G, et al., 2019. $\alpha$-Rank: multi-agent evaluation by evolution. *Sci Rep*, 9(1):9937. https://doi.org/10.1038/s41598-019-45619-9

Shao K, Zhu YH, Zhao DB, 2019. Starcraft micromanagement with reinforcement learning and curriculum transfer learning. *IEEE Trans Emerg Top Comput Intell*, 3(1):73-84. https://doi.org/10.1109/TETCI.2018.2823329

Shoham Y, Leyton-Brown K, 2008. Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511811654

Silver D, Huang A, Maddison CJ, et al., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484-489. https://doi.org/10.1038/nature16961

Smith JM, 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511806292

Von Neumann J, 1928. Zur theorie der gesellschaftsspiele. *Math Ann*, 100(1):295-320 (in German).

Von Neumann J, Morgenstern O, 2007. Theory of Games and Economic Behavior. Princeton University Press, Princeton, USA.

Wang MY, Li YM, Cheng ZX, et al., 2021. Evolution and equilibrium of a green technological innovation system: simulation of a tripartite game model. *J Clean Prod*, 278:123944. https://doi.org/10.1016/j.jclepro.2020.123944

Xu GZ, 2000. Systems Science and Engineering. Shanghai Scientific Technologically Education Publishing House, Shanghai, China (in Chinese).