RESEARCH ON FUSION ALGORITHM OF MULTI-ATTRIBUTE DECISION MAKING AND REINFORCEMENT LEARNING BASED ON INTUITIONISTIC FUZZY NUMBER IN WARGAME ENVIRONMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Intelligent games have seen an increasing interest within the research community on artificial intelligence . The article proposes an algorithm that combines the multi-attribute management and reinforcement learning methods, and that joined their effect on wargaming, it solves the problem of the agent's low rate of winning against specific rules and its inability to quickly converge during intelligent wargame training. At the same time, this paper studied a multi-attribute decision making and reinforcement learning algorithm in a wargame simulation environment, yielding data on the conflict between red and blue sides. We calculate the weight of each attribute based on the intuitionistic fuzzy number weight calculations. And then we determine the threat posed by each opponent's game agents . Using the red side reinforcement learning reward function, the AC framework is trained on the reward function, and an algorithm combining multi-attribute decision making with reinforcement learning is obtained. A simulation experiment confirms that the algorithm of multi-attribute decision making combined with reinforcement learning presented in this paper is significantly more intelligent than the pure reinforcement learning algorithm. By resolving the shortcomings of the agent's neural network, coupled with sparse rewards in large-map combat games, this robust algorithm effectively reduces the difficulties of convergence. It is also the first time in this field that an algorithm design for intelligent wargaming combines multi-attribute decision making with reinforcement learning. Finally, another novelty of this research is the interdisciplinary, like designing intelligent wargames and improving reinforcement learning algorithms. ABSTRACT must be centered, in small caps, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) are becoming increasingly popular in real-world applications. For example, AlphaGo has attracted huge attention in the research community and society by showing the capability of AI defeating professional human players in the board game Go. Yet Alphastar, another strong AI program, has achieved great success in the human-machine combating game 'StarCraft' Pang et al. (2019); Silver et al. (2016). In RTS games, AI-driven methods are widely studied and integrated into the game AI design to increase the intelligence of computer opponent and generate more realistic confrontation gaming experience. In the King Glory Game, Ye D used an improved PPO algorithm to train the game AI, with positive results Ye et al. (2020). By using reinforcement learning techniques, Silver D et al. developed a training framework that requires no human knowledge other than the rules of the game, allowing AlphaGo to train itself, and achieving high levels of intelligence in the process Silver et al. (2017). Using deep reinforcement learning and supervised strategy learning, Barrigan el al. improved the AI performance of RTS games, and defeats the built-in game AI Barriga et al. (2019). AI has become a hot research topic in recent years, showing a wide variety of applications such as deduction and analysis Schrittwieser et al. (2020); Barriga et al. (2017); O'Hanlon (2021). However, there are still limited research to

address the problem of slow convergence during AI training process under a variety of conditions, especially when it comes to human-AI confrontation games.

Indexes measure the value of things or the parameter of an evaluation system. It is the scale of the effectiveness of things to the subject. As an attribute value, it provides the subjective consciousness or the objective facts expressed in numbers or words. It is important to select a scientifically valid target threat assessment (TA) index and evaluate that index scientifically. Target threat assessment contributes to intelligence wargame decision-making as part of current intelligent wargames. It is mainly based on rules, decision trees, reinforcement learning, and other technologies in the current mainstream game intelligent decision-making field, but rarely incorporates multi-attribute decision-making theory and methods into the intelligent decision-making field. The actual wargame data obtained through wargame environments are presented in this paper, as well as the multi-attribute threat assessment indicators that are effectively transformed and presented as a unified expression. Using three expression forms of real number, interval number, and intuitionistic fuzzy number, the multi-attribute decision-making theory and methods are used to analyse the target threat degree. Then , an enhanced reward function based on the generated threat degree is established to train more effective intelligent decision making model. To the best of our knowledge, this is the first work that combines the multi-attribute decision making with reinforcement learning to produce high performance for game AI in a wargame experiment.

2 WARGAMING MULTIPLE ATTRIBUTE INDEX THREAT QUANTIFICATION

Obtaining scientific evaluation results requires a reasonable quantification of indicators. An important aspect of decision-making assistance in wargames is target threat assessment, and the evaluation result directly affects the effectiveness of wargame AI. The aim of this section is to introduce threat quantification methods for different types of indicators. By combining the target type, this section divides the target into target distance threat, target attack threat, target speed threat, terrain visibility threat, environmental indicator threat, and target defense value. The acquired confrontation data are incorporated into different indicator types, and then the corresponding comprehensive threat value is calculated. In Table 1 are the attributes and meanings of specific indicators.

Table 1: **A list of indicator attributes and their meanings**

Indicator	Attribute	Meaning
Target distance threat	Cost type	Distance between the two parties will influence the kill probability.
Target attack threat	Benefit type	Threat degrees should be
determined by the opponent's type, range, and lethality of the weapon.		
Target speed threat	Benefit type	The threat of speed from our opponents.
Terrain visibility threat	Intervisibility $>$ no intervisibility	Whether or not the terrain is visible will directly impact the threat.
Environmental indicator threat	Benefit type	While the opponent's environment
is conducive to concealment, mobility is more dangerous.		
Target defence value	Cost type	The stronger the opponent's armor, the harder it is to destroy it.

3 ESTABLISHMENT OF A MULTI-ATTRIBUTE QUANTITATIVE THREAT MODEL BASED ON INTUITIONISTIC FUZZY NUMBERS

By using the interval number method, our framework indicates whether visibility is possible, and different threats are generated. Nevertheless, the quantified values of other threat targets are real numbers. To unify the problem-solving method, our algorithm converts all interval numbers and real numbers to intuitionistic fuzzy numbers, and calculates the size of the threat by calculating the intuitionistic fuzzy numbers.

(1) This intuitionistic fuzzy entropy describes the degree of fuzzy judgment information provided by an intuitionistic fuzzy set. The larger the intuitionistic fuzzy entropy of an evaluation criterion, the smaller the weight it is; otherwise, the larger needs to be. Based on formulas from the literature Vlachos & Sergiadis (2007), we calculated the entropy weights for each intuitionistic fuzzy. Among them, ideal solution S_i^+ is a conceived optimal solution (scheme), and its attribute values hit the best value among the alternatives; and the negative ideal solution S_i^- is the worst conceived solution (scheme), and its attribute values hit the worst value among the alternatives. p_i is generated by comparing each alternative scheme with the ideal solution and negative ideal solution. If one of the solutions is closest to the ideal solution, but at the same time far from the negative ideal solution,

then it is the best solution among the alternatives.

$$H_j = -\frac{1}{n \ln 2} \sum_{i=1}^{m} [\mu_{ij} \ln \mu_{ij} + v_{ij} \ln v_{ij} - (\mu_{ij} + v_{ij}) \ln (\mu_{ij} + v_{ij}) - (1 - \mu_{ij} - v_{ij}) \ln 2] \quad (1)$$

If $\mu_{ij} = 0, v_{ij} = 0$, then $\mu_{ij} \ln \mu_{ij} = 0, v_{ij} \ln v_{ij} = 0, (\mu_{ij} + v_{ij}) \ln (\mu_{ij} + v_{ij}) = 0$.

The entropy weight of the j attribute is defined as:

$$w_j = \frac{1 - H_j}{n - \sum_{j=1}^{n} H_j} \quad (2)$$

Among $w_j \geq 0, j = 1, 2, \cdots, n, \sum_{j=1}^{n} w_j = 1$

(2) Determine the optimal solution A+ and the worst solution A- using the following formula:

$$\begin{cases} A^+ = \left\{ \left\langle \mu_1^+, \nu_1^+ \right\rangle, \left\langle \mu_2^+, \nu_2^+ \right\rangle, \cdots, \left\langle \mu_n^+, \nu_n^+ \right\rangle \right\} \\ A^- = \left\{ \left\langle \mu_1^-, \nu_1^- \right\rangle, \left\langle \mu_2^-, \nu_2^- \right\rangle, \cdots, \left\langle \mu_n^-, \nu_n^- \right\rangle \right\} \end{cases} \quad (3)$$

Where

$$\mu_i^+ = \max_{j=12.....m} \{\mu_{ij}\}, \nu_i^+ = \min_{j=1,2,...,m} \{\nu_{ij}\} \quad (4)$$

$$\mu_i^- = \min_{j=1,2,\cdots,m} \{\mu_{ij}\}, \nu_i^- = \max_{j=1,2,\cdots,m} \{\nu_{ij}\} \quad (5)$$

(3) Calculate the similarity between the fuzzy intuitionistic A and B as follows:

$$s(\langle \mu_1, \nu_1 \rangle, \langle \mu_2, \nu_2 \rangle) = 1 - \frac{|2(\mu_1 - \mu_2) - (\nu_1 - \nu_2)|}{3} \times \left(1 - \frac{\pi_1 + \pi_2}{2}\right) - \frac{|2(\nu_1 - \nu_2) - (\mu_1 - \mu_2)|}{3} \times \left(\frac{\pi_1 + \pi_2}{2}\right) \quad (6)$$

In which, $\pi_1 = 1 - \mu_1 - \nu_1, \pi_2 = 1 - \mu_2 - \nu_2$

(4) Calculate the similarity S_i^+ and S_i^- between each solution and the optimal solution and the worst solution based on the following formula:

$$\begin{cases} S_i^+ = \sum_{k=1}^{n} w_k \cdot s\left(\left\langle \mu_k^+, \nu_k^+ \right\rangle, \langle \mu_{ik}, \nu_{ik} \rangle\right) \\ S_i^- = \sum_{k=1}^{n} w_k \cdot s\left(\left\langle \mu_k^-, \nu_k^- \right\rangle, \langle \mu_{ik}, \nu_{ik} \rangle\right) \end{cases} \quad (7)$$

(5) Then calculate the relative closeness

$$p_i = S_i^- / \left(S_i^+ + S_i^-\right) \quad (8)$$

Comparing threat levels of opponents based on their closeness to the target depends on the level of threat assessment performed.

4 MULTI-ATTRIBUTE THREAT QUANTITATIVE SIMULATION

The threat assessment problem is transformed into a multi-attribute decision making problem, while the combat intention of the target is incorporated into the evaluation system to make the evaluation more realistic and the results more reliable. A simulation scene includes ten tanks on each side, i.e. red and blue, fighting each other, and ten opposite are found as game agents in the wargame.

A unified intuitiveistic fuzzy number representation has been created for all multi-attribute indicators. An example of an intuitionistic fuzzy number representation of threat assessment indicators is illustrated in Table 2.

Table 2: **Information decision table for threat target parameters (intuitionistic fuzzy number)**

	Tank1	Tank2	Tank3	Tank4	Tank5	Tank6	Tank7	Tank8	Tank9	Tank10
Quantification of target distance threats	[0.187378998, 0.012621002]	[0.18749387, 0.01250613]	[0.187608882, 0.012391118]	[0.176663586, 0.023336414]	[0.187608882, 0.012391118]	[0.176663586, 0.023336414]	[0.176561598, 0.023438402]	[0.199738767, 0.000261233]	[0.2, 0.0]	[0.2, 0.0]
Quantification of target speed threats	[0.153863899, 0.046136101]	[0.171440811, 0.028559189]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.171440811, 0.028559189]	[0.186672886, 0.013327114]	[0.171440811, 0.028559189]
Quantifying the threat from target attacks	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]
Quantifying the threat posed by terrain visibility	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]
Quantification of environmental indicators of threat	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[0.2, 0.0]	[6.6644e-05, 0.199933356]	[0.2, 0.0]
Quantification of target defense	[0.2, 0.0]	[0.2, 0.0]	[0.000664452, 0.199335548]	[0.000399202, 0.199600798]	[0.0001998, 0.1998002]	[0.2, 0.0]	[0.2, 0.0]	[0.00664452, 0.199335548]	[0.0001998, 0.1998002]	[0.000285307, 0.199714693]

Table 3: **Threat assessment for target**

S_i^+	[0.9900131572106283, 0.9930194457658972, 0.9713249517102417, 0.9694274902547305, 0.9712630240082707, 0.9960298049584839, 0.9960124538670997, 0.9685356920167532, 0.9447732710194203, 0.9685296037271114]
S_i^-	[0.9451975215527424, 0.9421912329974735, 0.9638857270531 29, 0.9657831885086402, 0.9639476547551001, 0.9391808738048868, 0.9391982248962711, 0.9666749867466174, 0.9904374077439504, 0.9666810750362593]
P_i	[0.511579006913739 1, 0.5131324752716081, 0.5019220710020746, 0.5009415775207532, 0.5018900705058751, 0.5146880470889931, 0.5146790810929003, 0.5004807500523212, 0.4882017660336315, 0.5004776039919 42]
Ranking	T6>T7>T2>T1>T3>T5>T4>T8>T10>T9

By obtaining data represented by the intuitionistic vagueness of the threat assessment indicators shown in the Table 2, formulae in (7) and (8) may be used to obtain the intuitionistic vague target threat assessment based on multi-attribute decision making approaches. Table 3 shows the assessment scores to determine the target threat level.

In Table 4, the opposite target at $T1$ is shown as a threat.

Table 4: **Ranking of opposite targets at time Tt**

Type of piece	Indicator comprehensive	Ranking
Tank 1	0.511579007	4
Tank 2	0.513132475	3
Tank 3	0.501922071	5
Tank 4	0.500941578	7
Tank 5	0.501890071	6
Tank 6	0.514688047	1
Tank 7	0.514679081	2
Tank 8	0.50048075	8
Tank 9	0.488201766	10
Tank 10	0.500477604	9

Based on the evaluation results, it can be concluded that the blue $T6$ tank is the most harmful and the $T7$ tank is the second most harmful, this is shown in figure 1. This paper does not limit evaluation to subjective analysis of experts, but also introduces reinforcement learning, associates the reinforcement learning algorithm through a reward function and analyses the actual wargame AI's winning rate.

5 A Fusion Model of Reinforcement learning and multi-attribute threat analysis

5.1 Reinforcement learning algorithm and multi-attribute model formulation

Previous sections described the quantified value of multi-attribute analysis of threat levels based on the entropy weight method. The section integrate this method with with reinforcement learning. Its essence is to establish a multi-attribute decision-making mechanism that is based on reinforcement learning, and then select the entity with the highest threat level to establish the return value and threat level. The higher the threat level, the greater the return value, this is shown in figure 2.

A reinforcement learning algorithm is built using the AC framework to achieve intelligent decision-making. It includes a reinforcement learning pre-training module that integrates multi-attribute decision-making, critic evaluation network update module and a new and old strategy network update module. In the intensive pre-training module, multi-attribute decision making mainly uses state data obtained from the wargame environment, such as elevation, distance, armour thickness, etc., to make multi-attribute decisions. By normalizing the data, calculating the threat of each piece of the opponent by using the entropy method, and then setting the reward function and storing it in the experience, further actions in the environment will be taken to obtain the next state and action rewards.

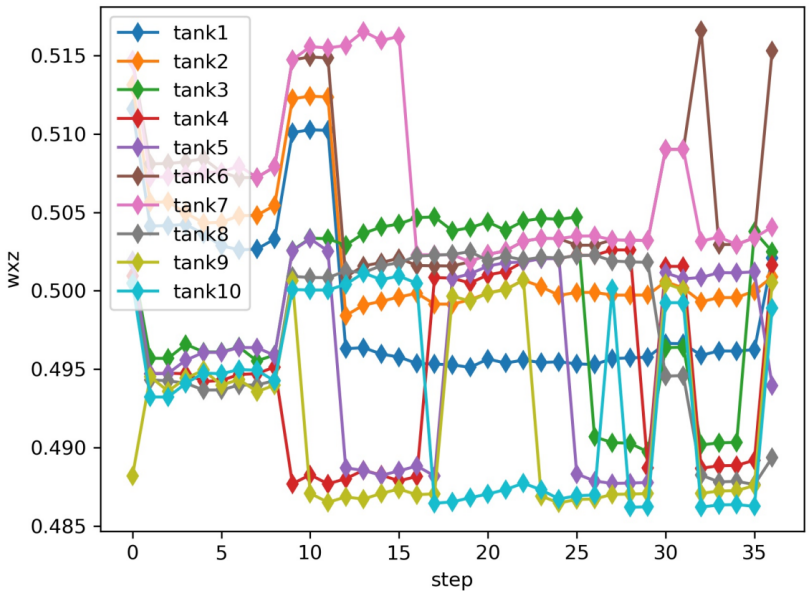


Figure 1: The threat value on the ordinate, and the threat of the opponent's ten tanks at time T represented by ten colours on the abscissa.

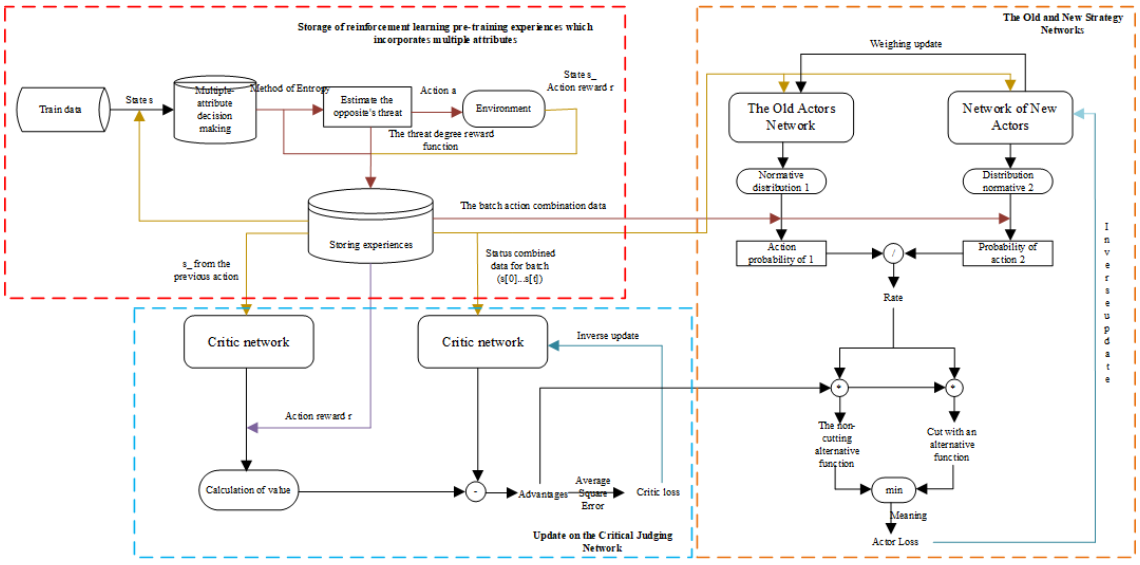


Figure 2: A fusion model of reinforcement learning and multi-attribute threat estimation based on AC framework. The module mainly consists of a reinforcement learning pre-training module that integrates multi-attribute decision-making, Critic evaluation network update module, and a new and old strategy network module

The critic network calculates the value from the reward value determined during the last step of the action. combines the experience store data with the value calculated by the critic network, slashes it from the reward value determined during the last action, then returns to update the critic network parameters. As the advantage value guides the calculation of the actor network value, the network outputs the action value according to the old and new networks, and the distribution probability overall, and outputs the action from the network. As a result, the advantage value is corrected, the actor loss is calculated, and the actor network is updated in the reverse direction.

5.2 SETTING REWARD FUNCTION VALUE

As a core challenge of deep reinforcement learning in solving practical tasks, the sparse reward problem relates to the fact that the training environment cannot supervise the updating of agent parameters in the process of reinforcement learning Kaelbling et al. (1996). When supervised learning is used, the training process is supervised by humans, while in reinforcement learning, rewards are used to supervise the training process, and the agent optimizes strategies based on rewards **?**. The specific additional rewards is showed in Table 5.

Table 5: **Reward settings**

Situation	Reward
The state is now closer to the control point than the previous state	Reward+0.5
This state is nearly as far from the control point as the previous state	Reward-0.3
The map boundary has been reached	Reward-1
Consumption per step (to avoid falling into local optimum)	Reward-0.005
The opposite piece was hit	Reward+(5*Risk of being hit by a piece)
Hit by an opposite round	Reward-(5*Risk of being hit by a piece)
An opposite piece is annihilated	Reward+10
Taking out one of the opposite's pieces will lead to victory	Reward+20
Defeat an opposite piece leading to failure (other opposite pieces reach the control point)	Reward-10
Get to the control point	Reward+10
opposite wins	Reward-10

When the above additional rewards are added to the training process, the convergence speed can be significantly accelerated, and the likelihood that the agent falls into the local optimum is significantly reduced.

6 WARGAMES AI SIMULATIONS AND EVALUATIONS

6.1 EXPERIMENT SETTING

Figure 3 shows the starting interface of our simulation which generates the initial states of red and blue tanks Sun et al. (2021) Sun et al. (2020). There are two tank pawns on each side, and the centre is the point of contention. In a confrontation, both sides compete for control points, and the party that reaches the middle red flag first wins. At the same time, both red and blue parties can shoot at each other, while they can hide in urban residential areas. By concealing, it is difficult for our opponents to find our targets. Each hexagon has its own number and elevation. The higher the elevation, the darker the hexagon. On the highway, the tanks move faster than on the secondary roads. The red straight line represents the secondary road and the black straight line represents the primary road. As the cross symbol represents aiming and shooting, the destroyed target disappears from the map.

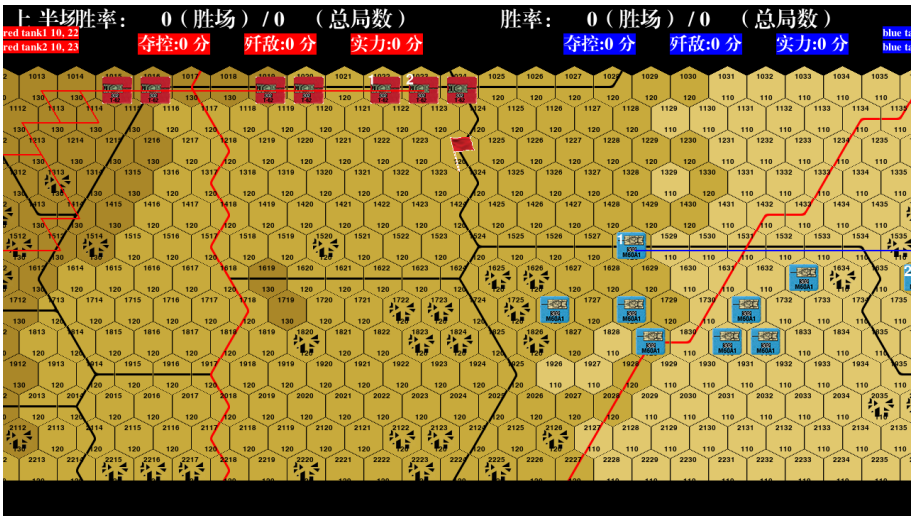
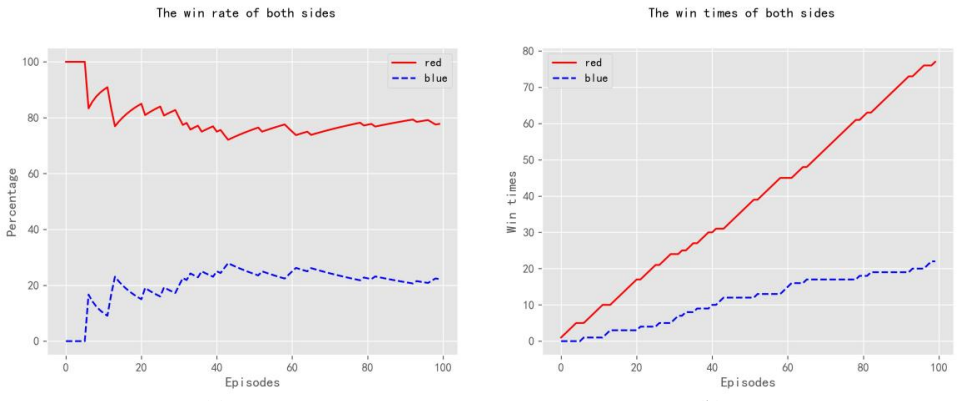


Figure 3: Gaming environment display. The red and blue pawns fight separately, the red flag in the middle is the control point, and the first player to reach the control point wins. Alternatively, when all the wargame agents on one side are destroyed, the opponent wins.

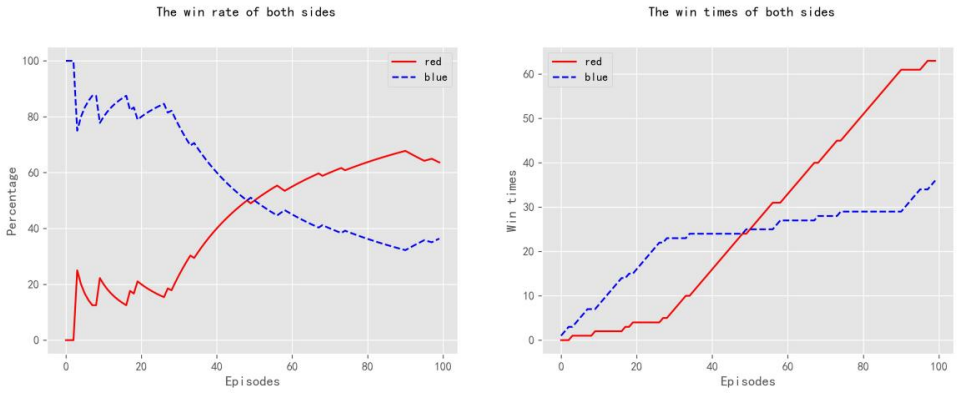
6.2 Results and Analysis of the Experiment

In this article, the PPO algorithm Schulman et al. (2017) and the PPO algorithm combined with multi-attribute decision-making are used to compare and analyse the winning rate. MADM-PPO and PPO are trained for 24 hours, and this article uses the MADM-PPO algorithm as the red side and the rule-based blue side algorithm to fight. At the same time, the second round uses the PPO algorithm as the red side, and the blue side fights according to rules. Next, this article observes the winning percentage of both algorithms in 100 games. Experiments have shown that the agents using the PPO reinforcement learning algorithm combined with the multi-attribute decision-making method performed better than the agents using the PPO algorithm based on the threat of the opponent. As can be seen in the Figure 4 and Figure 5, our proposed multi-attribute decision-making method, combined with PPO algorithm of reinforcement learning, proves to effectively improve the effectiveness of intelligent wargame decision-making. A winning rate chart is presented in the Table 6, and Table 7.



(a) (b)

Figure 4: (a) Win rate: the red side is the AI of MADM-PPO intelligent algorithm and the blue side is rule-based AI; (b) Win times: the red side is the AI of MADM-PPO intelligent algorithm and the blue side is rule-based AI; The winning rate and the number of wins for the red and blue sides. The first round wins so one side starts from 1 and the other from 0.



(a) (b)

Figure 5: (a) Win rate: the red side is the AI of PPO intelligent algorithm and the blue side is rule-based AI; (b) Win times: the red side is the AI of PPO intelligent algorithm and the blue side is rule-based AI; The winning rate and the number of wins for the red and blue sides. The first round wins so one side starts from 1 and the other from 0.

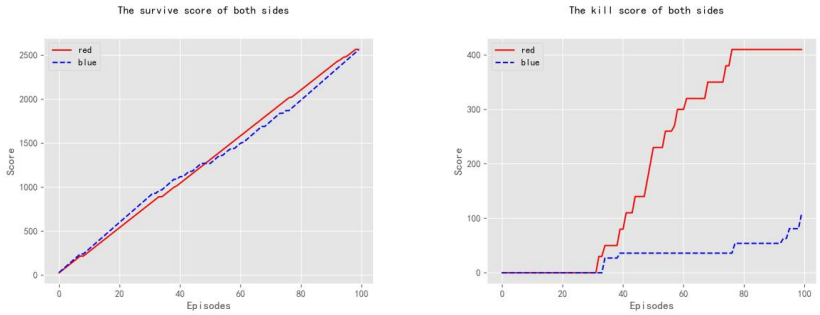
The experimental results show that the MADM-PPO model can reduce the number of times to explore during training, and improve the problem that the PPO algorithm takes too long to train. It shows that the introduction of prior knowledge improves the performance of the PPO algorithm, and has a certain theoretical significance for improving the efficiency of the algorithm, the detail score is shown in Figure 6.

7 Conclusion

We have designed an intelligent wargaming AI that To design intelligent wargaming AI that combines multi-attribute decision making and reinforcement learning to improve both the convergence speed of the online training process and the winning rate of wargaming AI. As part of this study, this paper conducts experiments on the multi-attribute decision making and reinforcement learning algorithms in a wargame simulation environment, and obtains red and blue confrontation data from the wargame environment. Calculate the weight of each attribute based on the intuitionistic fuzzy number weight calculations. Then determine the threat posed by each opponent's game agents . On the basis of the degree of threat, the red side reinforcement learning reward function is constructed and the AC framework is trained with the reward function, and the algorithm combines multi-attribute decision making with reinforcement learning. A study demonstrated that the algorithm can gradually increase the reward value of the agent when exploring an environment over a short training period, while the final victory rate of the agent against specific rules and strategies reached 78%, which is significantly higher than that of a pure reinforcement learning algorithm, which is 62%. Solved the convergence difficulties of the state-space wargame's sparse rewards caused by the randomization of an agent's neural network. For the algorithm design of intelligent wargaming, this is the first research in this field to combine the multi-attribute decision making method in management with the reinforcement learning algorithm in cybernetics. An interdisciplinary approach to cross-innovation in academia could lead to improvements in the design of intelligent wargames and even improvements in reinforcement learning algorithms. The future research direction can be based on this paper to carry out a series of research, including the introduction of new methods in management multi-attribute decision-making and the fusion and intersection of a series of algorithms such as reinforcement learning SAC, MADDPG and DDQN etc, which can develop more, better and more stable fusion innovative algorithms.

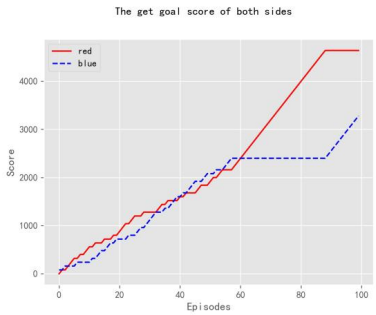
References

Nicolas A Barriga, Marius Stanescu, and Michael Buro. Combining strategic learning with tactical search in real-time strategy games. In *Thirteenth Artificial Intelligence and Interactive Digital*

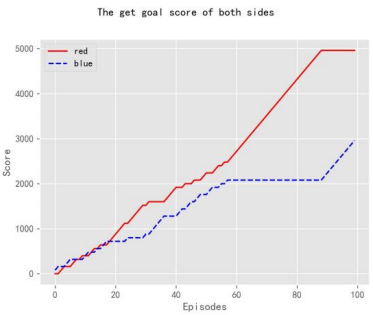


(a)

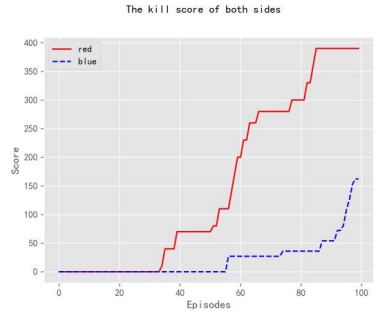
(b)



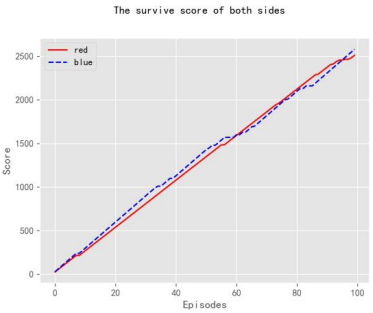
(c)



(d)



(e)



(f)

Figure 6: (a) The get goal score of both sides (Red: PPO); (b) the kill score of both sides (Red: PPO); (c) the survive score of both sides (Red: PPO); (d) the get goal score of both sides (Red: MADM-PPO); (e) the kill score of both sides (Red: MADM-PPO); (f) the survive score of both sides (Red: MADM-PPO). The x-axis is the training episodes, and the y-axis is the score. Red and blue represent two teams in the wargame environment.

Entertainment Conference, 2017.

Nicolas A Barriga, Marius Stanescu, Felipe Besoain, and Michael Buro. Improving rts game ai by supervised policy learning, tactical search, and deep reinforcement learning. *IEEE Computational Intelligence Magazine*, 14(3):8–18, 2019.

Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

Michael E O'Hanlon. 2. gaming and modeling combat. In *Defense 101*, pp. 85–133. Cornell University Press, 2021.

Zhen-Jia Pang, Ruo-Ze Liu, Zhou-Yu Meng, Yi Zhang, Yang Yu, and Tong Lu. On reinforcement learning for full-length game of starcraft. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4691–4698, 2019.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

Yuxiang Sun, Bo Yuan, Tao Zhang, Bojian Tang, Wanwen Zheng, and Xianzhong Zhou. Research and implementation of intelligent decision based on a priori knowledge and dqn algorithms in wargame environment. *Electronics*, 9(10):1668, 2020.

Yuxiang Sun, Bo Yuan, Yongliang Zhang, Wanwen Zheng, Qingfeng Xia, Bojian Tang, and Xianzhong Zhou. Research on action strategies and simulations of drl and mcts-based intelligent round game. *International Journal of Control, Automation and Systems*, pp. 1–15, 2021.

Ioannis K Vlachos and George D Sergiadis. Intuitionistic fuzzy information–applications to pattern recognition. *Pattern Recognition Letters*, 28(2):197–206, 2007.

Deheng Ye, Zhao Liu, Mingfei Sun, Bei Shi, Peilin Zhao, Hao Wu, Hongsheng Yu, Shaojie Yang, Xipeng Wu, Qingwei Guo, et al. Mastering complex control in moba games with deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 6672–6679, 2020.