Research on Performance Evaluation of Wargame System Based on Deep Reinforcement Learning

Tongfei Shang¹,², Jinyun Wu¹ and Jianfeng Ma¹

¹ College of information and communication, national university of defense technology, Xi’an, China.
² Department of Computer Science and Technology, School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, China.
Email: 340445698@qq.com

Abstract. Deep reinforcement learning combines the advantages of deep learning and reinforcement learning to make end-to-end perception decisions in complex high-dimensional state action spaces. The paper proposes a deep reinforcement learning method for the deduction of the wargame system, which can better assist the combat commander in wartime decision-making. The simulation proves the effectiveness of the method.

1. Introduction

In the early stage of deep learning, multi-layer perceptions were proposed and back-propagation algorithms were used to optimize multi-layer neural networks. However, due to the problems of gradient dispersion or explosion and the limitation of hardware resources, no breakthrough progress was made. Reinforcement learning interacts with the environment in a trial-and-error mechanism to learn optimal strategies by maximizing cumulative rewards. In this paper, the deep learning and reinforcement learning are combined with the advantages and disadvantages of the two learning, and a new machine learning algorithm is generated. The simulation results show that the algorithm can effectively carry out the force distribution of the wargame system and better evaluate the derivation performance of the war chess system. It has an important role.

2. Deep Reinforcement Learning

Deep reinforcement learning is a new topic in the field of machine learning. The goal is to learn a control strategy to maximize the expected cumulative rewards of all states (or “state-action” pairs) and to be an effective means of solving sequential decision problems. The use of deep reinforcement learning for war chess deduction will greatly improve the intelligence of the assistant decision-making system and better distribute the firepower.

Given a sequence, a set of time series X, a distance metric D, and a threshold value, the problem of time series similarity search is to find a sequence R with a distance D less than the threshold, that is,

\[ R = \{ \tilde{x} \in X \mid D(\tilde{q}, \tilde{x}) \leq \varepsilon \} \] (1)

The threshold value is generally given by the user. The specific form of the distance function is related to the problem to be solved and the algorithm used. The Euclidean space distance between X and Y is often used instead. However, the spatial distance does not allow transformations in the form of translation, scaling, etc. on the time axis as well as on the numerical axis. According to the definition of the rising and falling trend of time series data in the intuitive sense, the timing with...
similar shape can be found by this similarity definition; let $F$ be a function set, if they are satisfied for the timings to be compared a sub-sequence of a certain length, and a function $f$ in $F$ such that one of the sub-sequences can be approximated to another sub-sequence, the two timings are said to have $F$-similarity. This similarity is insensitive to outliers, baselines, and scale factors in time series; element values are generally continuous and elements are not necessarily uniformly distributed, etc. These characteristics make it difficult to directly apply the game. The original algorithm is generally only applicable to the case of low dimensionality, but to turn the time series directly into a vector. 

3. Wargame System Deduction Performance Evaluation Wavelet Analysis

The deep reinforcement learning algorithm initially adopted simple exploration strategies such as ε-greedy strategy and Boltzmann exploration strategy. The current exploration strategy based on the knowledge of war chess is based on the current situation of fuzzy (qualitative) rules. This paper proposes a deep reinforcement learning exploration strategy, which expresses qualitative knowledge based on mathematical models and guides the action selection of reinforcement learning.

If the measured data in the pawn system is set, and the model output data is $\mathbf{y}$, then the output error of the starting model in each pawn deduction is defined as:

$$E(N1,) = \frac{1}{T} \sum_{i=1}^{15} |N1h_{i}^{'} - N1h_{i}|, \quad i = 1,2,\ldots,185$$

(2)

$$E(N2,) = \frac{1}{T} \sum_{i=1}^{15} |N2h_{i}^{'} - N2h_{i}|, \quad i = 1,2,\ldots,185$$

(3)

$$E(T4,) = \frac{1}{T} \sum_{i=1}^{15} |T4_{i}^{'} - T4_{i}|, \quad i = 1,2,\ldots,185$$

(4)

The output error of the wargame system deduction evaluation model is defined as:

$$E(N1,) = \frac{1}{T} \sum_{i=1}^{15} |N1h_{i}^{'} - N1h_{i}|, \quad i = 1,2,\ldots,185$$

(5)

$$E(N2,) = \frac{1}{T} \sum_{i=1}^{15} |N2h_{i}^{'} - N2h_{i}|, \quad i = 1,2,\ldots,185$$

(6)

The output error of the evaluation model is defined as:

$$E(T4,) = \frac{1}{15} \sum_{i=1}^{15} |T4_{i}^{'} - T4_{i}|, \quad i = 1,2,\ldots,185$$

(7)

In equations (2) to (7), the duration of the deduction process for the siege system is the order number. According to the structural framework of the war chess system, a prototype system for evaluating the war chess data is realized, and the prototype system is integrated into the computer war chess command release system, so that the performance of the war chess system can be evaluated in real time during the deduction process.

4. Simulation Analysis

If the output error of each parameter in the above model is arranged into a sequence according to the order of the derivation data of the war chess system, the deviation (the third type of error) due to the deviation of the actual model from the initial model is included in the set of errors. In the sequence, the magnitude of the deviation reflects the degree of evaluation of the derivation performance of the war chess system relative to the initial state. In order to reduce the influence of the derivation evaluation (the first type of error), the obtained error sequence is multi-scale-decomposed, and the function is bi-orthogonal and partially reconstructed on the scale.
In Fig. 1, the first of each set of curves is the original error sequence, and the following are the reconstruction curves at different scales. The error of the first 20 points is very small, and then the overall error of the derivation performance of the war chess system is increasing due to the gradual increase of the third type of error (deviation). The original data curve fluctuates very sharply, and the reconstruction curve on the scale and the upper is slightly flat, but there are still small fluctuations. The reconstruction curve on the scale can clearly see that there is no influence of the deduced packet data, and it retains the growth with time. Increased amount of deviation.

In this paper, based on the calculation of the W value of each parameter in the three typical working stages of the performance of the game, the average value of W is counted, as shown in Table 1. In Tab. 1, W1-W4 is the deviation value of each parameter of the derivation performance of the chess game system.

![Figure 1. Wargame system derivation performance evaluation bias](image1)

![Figure 2. Wargame system derivation performance evaluation bias](image2)
Table 1. Wargame system derivation performance evaluation difference

| W1 | N1hs (%) | N2hs (%) | T4 (s) | W2 | N1hs (%) | N2hs (%) | T4 (s) | W3 | N1hs (%) | N2hs (%) | T4 (s) | W4 | N1hs (%) | N2hs (%) | T4 (s) |
|----|---------|---------|-------|----|---------|---------|-------|----|---------|---------|-------|----|---------|---------|-------|
| W1 | 0.126   | 0.155   | 1.861 | W2 | 0.243   | 0.392   | 4.395 | W3 | 0.381   | 0.887   | 6.628 | W4 | 0.577   | 1.212   | 10.852 |
| W1 | 0.097   | 0.201   | 2.825 | W2 | 0.223   | 0.536   | 7.638 | W3 | 0.494   | 0.858   | 11.257| W4 | 0.725   | 1.145   | 18.684 |

5. Conclusion
Aiming at the characteristics of computer wargame performance, this paper proposes a method based on deep reinforcement learning to evaluate the performance of wargame system. The experimental results show that deep reinforcement learning can greatly shorten the estimation error of the deduction in the case of large-scale forces between the enemy and the enemy. Deep reinforcement learning research has great application value for training guidance and post-evaluation in computer wargame system, and it also has a guiding role for artificial intelligence in strategy games.

6. References
[1] Preston DR. Brodley C E, Khardon R. Redefining class definitions using constraint—based clustering[C]. Proc. of the 16th ACM S, GKDD Conference on Knowledge Discovery and Data Mining. 2010: 823-832.
[2] Liu N, Zhang B Y, Yan J. et al. Learning similarity measures in the nonorthogonal space[C]}.Proc. of the 13th Conference on Information and Knowledge Management. 2004:334-341.
[3] CUI Yan. SONG Yexin. Hypergame stability analysis fuzzy preference perception marx[J]. Joumal of Naval University of Engineering, 2007, 19(6): 103—107.