融合多属性决策和深度Q值网络的反导火力分配方法

谢俊伟[①] 方 峰[*①] 彭冬亮[①] 任金磊[②] 王昌平[①]

[①](杭州电子科技大学自动化学院 杭州 310018)

[②](中国运载火箭技术研究院 北京 100076)

摘 要：针对中大规模武器-目标分配(WTA)决策空间复杂度高、求解效率低的问题，该文提出一种基于多属性决策和深度Q网络(DQN)的WTA优化方法。建立基于层次分析法(AHP)的导弹威胁评估模型，引入熵值法表征目标属性差异，提升威胁评估客观性。根据最大毁伤概率准则，建立基于DQN的WTA分段决策模型，引入经验池均匀采样策略，确保各类目标分配经验的等概率抽取；设计综合局部和全局收益的奖励函数，兼顾DQN火力分配模型的训练效率和决策准确性。仿真结果表明，相较于传统启发式方法，该方法具备在线快速求解大规模WTA问题的优势，且对于WTA场景要素变化具有较好的鲁棒性。

关键词：火力分配；深度Q网络；威胁评估；改进层次分析法

中图分类号：TP183; TJ761.7 **文献标识码**：A **文章编号**：1009-5896(2022)11-3833-09

DOI: 10.11999/JEIT211136

Weapon-Target Assignment Optimization Based on Multi-attribute Decision-making and Deep Q-Network for Missile Defense System

XIE Junwei[①] FANG Feng[①] PENG Dongliang[①]

REN Jinlei[②] WANG Changping[①]

[①](*School of Automation*, *Hangzhou Dianzi University*, *Hangzhou* 310018, *China*)

[②](*China Academy of Launch Vehicle Technology*, *Beijing* 100076, *China*)

Abstract: In a large-scale Weapon-Target Assignment (WTA) problem, the explored solution space becomes enormous due to the curse of dimensionality, and it causes low-efficiency in searching optimization solution. For solving this problem effectively, a WTA optimization approach based on multi-attribute decision-making and Deep Q-Network (DQN) is proposed. Firstly, a threat-assessment model for attacking missiles is built based on the approach of Analytic Hierarchy Process (AHP). Meanwhile, an entropy method, used for evaluating the differences of target attributes, is introduced, to increase objective in computing threat-assessment results. Then, an assignment criterion of maximum intercept probability is designed based on assess results, and a multi-steps WTA decision model is built in DQN frame. A uniform experience sampling strategy is designed, making sure that each target type of assignment experience has the same probability to be selected. Furthermore, for balancing the DQN convergence speed and global optimum, a reward function that combines local and global rewards is designed. Lastly, simulation results shows that the proposed WTA approach has the advantage in solving large-scale WTA problem fast and effectively, compared with the general heuristic approach. Also, it presents the robust performance for WTA scenario elements variation.

Key words: Weapon Target Assignment(WTA); Deep Q Network(DQN); Threat assessment; Improved Analytic Hierarchy Process(AHP)

收稿日期：2021-10-15；改回日期：2022-01-10；网络出版：2022-02-02

*通信作者： 方峰 fangf@hdu.edu.cn

基金项目：国家自然科学基金 (61673146)，浙江省属高校科研基金(GK209907299001-021)

Foundation Items: The National Natural Science Foundation of China(61673146), Zhejiang Provincial University Research Foundation (GK209907299001-021)

1 引言

为了应对弹道导弹和高超声速飞行器等目标的威胁，各国相继发展了由预警探测系统、导弹拦截系统、指挥控制作战管理系统组成的全球一体化反导防御体系。武器-目标分配(Weapon-Target Assignment, WTA)是导弹防御系统中的核心决策内容，决策人员根据来袭导弹目标的威胁程度和防御系统的拦截弹资源配置情况，按照特定的火力打击策略，生成火力分配方案，最大限度上发挥防御系统的作战性能[1]。

WTA问题可以分解为WTA模型构建和WTA优化方法两部分。由于拦截空域会出现多个来袭目标，因此在建立WTA模型时，首先需要评估来袭目标的威胁程度，确定拦截优先级，并基于此设计多约束条件下的火力分配准则函数。目前，已有的威胁评估方法主要包括层次分析(Analytic Hierarchy Process, AHP)方法[2]、优劣解距离(Technique for Order Preference by Similarity to an Ideal Solution, TOPSIS)方法[3]、贝叶斯网络方法[4]、粗糙集方法[5]等。其中，AHP方法在构建指标权重判别矩阵时较为依赖主观经验；TOPSIS方法的指标信息熵计算对数据噪声较为敏感，从而影响评估准确性；贝叶斯网络模型结构的确定缺乏客观设计标准；基于粗糙集理论的方法当历史数据集规模较小时，存在评估规则难以准确提取的问题。由此，本文针对AHP方法计算指标权重较为主观的问题，引入了表征目标特性信息的熵值法来增加准则层指标权重确定的客观性，从而提升目标威胁评估的准确性。改进的AHP方法计算量小，实时性好，便于工程上实现。

WTA优化方法是指在WTA模型基础上建立快速高效的优化搜索算法，给出最优或者次优的火力分配方案。WTA优化问题实质上是一类整数型非线性组合优化问题，属于NP完全(NP-Complete)问题[6]。目前，已有的WTA优化方法包括分支定界法[7]、动态规划[8]、遗传算法[9]和粒子群算法[10]等，但是，上述方法在面对中大规模WTA问题时求解效率较低。分支定界和动态规划存在搜索空间维数爆炸问题，启发式算法搜索速度慢且容易陷入局部最优。基于强化学习的决策方法可避免以上问题，近年来已被广泛应用在棋类博弈[11]、机器人路径规划[12]及自主空战决策[13]等场景中。本文将强化学习方法引入到火力分配问题中，把WTA问题转化为一个多步决策问题。文献[14]采用强化学习算法解决反舰导弹火力分配问题，但仅将单步决策带来的毁伤概率增量作为奖励函数，火力分配决策的全局最优性很难保证，求解方案不够理想。另外，文献[14]的状态向量和动作向量设计不够灵活，使得训练所得的智能体难以应对场景参数变化的情况。本文在深度Q网络(Deep Q-Network, DQN)框架下建立了高效的火力分配方法：基于最大毁伤概率准则设计了兼顾快速收敛和全局收益的奖励函数，构建了火力单元状态集、目标库和经验池，并引入了公平采样策略，确保等概率学习各目标分配经验。大量仿真结果表明，本文所提改进AHP方法通过目标属性值分布差异可以更加客观地评估目标威胁度，DQN火力分配方法则可以根据目标导弹的威胁度和拦截弹的毁伤能力，快速求解中大规模WTA问题的拦截弹-目标分配方案，实现最大概率毁伤来袭目标群；同时，本文训练得到的DQN智能火力分配模型对包括目标-火力单元类型和数量、拦截弹毁伤概率等WTA场景参数变化具有一定的鲁棒性。

2 WTA问题描述

本文分别围绕目标威胁评估和WTA优化这两部分开展WTA问题研究。目标威胁评估指的是导弹防御系统对来袭目标进行预警探测、识别与跟踪，确定来袭目标的数量、种类以及相应的运动状态信息，并应用上述目标信息评估目标威胁度。其中，需要提取能反映目标特性差异的关键因素作为威胁度评估指标集，由此计算来袭目标的威胁度。对于导弹防御系统而言，不同目标的威胁度会引起拦截优先级的差异，且是WTA模型的关键参数，对于后续火力分配决策起着决定性的作用。

假设红方来袭目标类型数量为k，目标数量为$n=\sum_{i=1}^{k}n_i$，其中n_i为第i类目标的数量；蓝方导弹防御系统的拦截弹种类为l，火力单元(拦截弹)数量为$m=\sum_{i=1}^{l}m_i$，其中m_i为第i类火力单元的数量。令$[x_{ij}]_{m\times n}$为火力分配决策矩阵，其中该决策矩阵的行和列分别是拦截弹和目标按照$\{m_1,m_2,\cdots,m_l\}$和$\{n_1,n_2,\cdots,n_k\}$的类别顺序进行排列的，$x_{ij}=1$表示将第i个火力单元分配给第j个目标，$x_{ij}=0$则表示不分配。因此，WTA模型可以描述为

$$\begin{aligned}&\max_{x_{ij}}\ J(x_{ij})=\sum_{j=1}^{n}v_j\left(1-\prod_{i=1}^{m}(1-p_{ij})^{x_{ij}}\right),\\&\text{s.t.}\ \sum_{j=1}^{n}x_{ij}\le 1\,,\quad \sum_{i=1}^{m}x_{ij}\ge 1\end{aligned}\tag{1}$$

其中，v_j为由威胁评估方法得到的目标威胁值，p_{ij}为第i个火力单元对j个目标的毁伤概率，不等式

约束则分别表示每个火力单元最多只能分配1个目标，每个目标可以分配多个火力单元。

3　目标威胁评估

3.1　威胁评估因素定量分析

本文考虑4类典型目标，分别为近、中、远程弹道导弹和高超声速飞行器，导弹防御系统则考虑低、中和高层3类典型拦截弹，如分别由美国的爱国者拦截弹(Patriot Advanced Capability-3, PAC-3)、海基拦截弹(Standard Missile, SM-3)和陆基拦截弹(Ground-Based Interceptor, GBI)构成的低中高层导弹防御系统。根据弹道导弹和高超声速飞行器等目标的运动特性和固有属性，构造如下威胁评估指标：来袭目标攻击区域重要程度、目标剩余飞行时间、目标最大飞行高度、目标关机点速度和雷达反射面积(Radar Cross-Section, RCS)。其中，目标打击区域重要程度根据该区域的军事、政治、经济等影响力由上级指挥专家打分给出，对应的威胁度值可以量化为

$$\omega(r)=1-0.1I_j,\quad 1\le I_j\le 9 \tag{2}$$

其中，I_j为整数，代表第j个目标攻击区域的重要程度。目标剩余飞行时间越小，留给防御系统的反应时间越短，对应的威胁度越大。本文涉及的弹道导弹和高超声速飞行器的最大飞行高度区间差别较大，分别为200～1400 km和20～80 km(临近空间)，在相应的高度范围内，最大飞行高度越大则威胁程度越大。目标的关机点速度决定了目标的再入速度和攻击威力，关机点速度越大，则拦截窗口时间越短，较难拦截，目标的威胁程度也越大。目标的雷达反射面积越小，防御系统也越难跟踪，其威胁程度越大。结合上述分析，可分别建立各威胁指标对应的分段量化函数，以最大高度为例，其威胁指标量化函数可以描述为

$$\omega(h)=\begin{cases}\dfrac{1}{9}, & h\le 20\ \text{km}\\ \dfrac{8}{9}, & 20\ \text{km}<h\le 80\ \text{km}\\ \dfrac{1}{9}+\dfrac{1}{3}\times\dfrac{h-80}{200-80}, & 80\ \text{km}<h\le 200\ \text{km}\\ \dfrac{4}{9}+\dfrac{1}{3}\times\dfrac{h-200}{1000-200}, & 200\ \text{km}<h\le 1000\ \text{km}\\ \dfrac{7}{9}+\dfrac{1}{9}\times\dfrac{h-1000}{1200-1000}, & 1000\ \text{km}<h\le 1200\ \text{km}\\ \dfrac{8}{9}+\dfrac{1}{9}\times\max\left(\dfrac{h-1200}{1600-1200},1\right), & h\ge 1200\ \text{km}\end{cases} \tag{3}$$

综上，根据威胁指标量化函数，可以得到各来袭导弹目标的威胁因子评估向量。

3.2　基于熵值法的改进AHP

AHP将复杂的评估系统模型层次化，通过逐层比较各种评估因素的重要性进行评估分析[2]。在导弹威胁评估问题中，目标层为目标威胁评估值，准则层为威胁评估因素，方案层为待评估的目标弹。本文在准则层中引入熵值法，通过评估目标(来袭导弹)的指标属性信息熵来修正准则层指标权重的计算，提升指标权重判定的客观性。引入熵值法的改进AHP方法整体框架如图1所示，具体执行步骤如下：

步骤1　应用AHP方法计算准则层的指标权重。根据专家意见采用1～9标度法构建准则层(各威胁评估因素)的判别矩阵$\boldsymbol{A}$，则AHP方法下的指标权重向量$\boldsymbol{w}^{\text{AHP}}$可计算为

$$\left.\begin{aligned}&\lambda_{\max}\boldsymbol{w}_{\max}=\boldsymbol{A}\boldsymbol{w}_{\max}\\ &w_j^{\text{AHP}}=w_{\max,j}\Big/\sum_{j=1}^{5}w_{\max,j}\end{aligned}\right\} \tag{4}$$

其中，$\lambda_{\max}$为判别矩阵$\boldsymbol{A}$的最大特征值，$\boldsymbol{w}_{\max}$为对

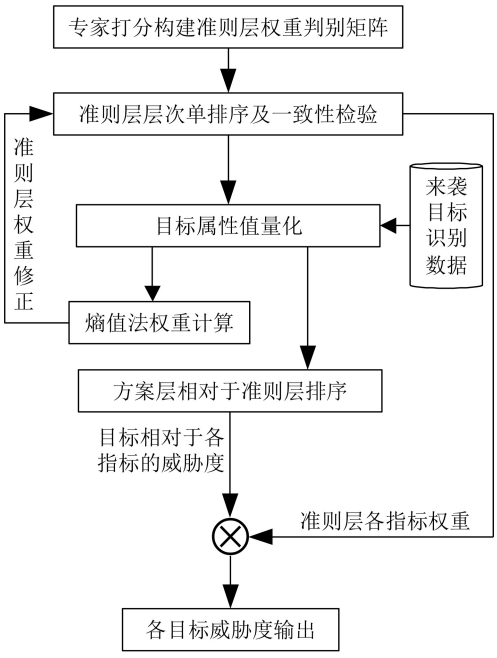

图 1 改进AHP法框架示意图

应的特征向量，$w_{\max,j}$ 为特征向量$\boldsymbol{w}_{\max}$中的第j个元素，w_j^{AHP}为权重向量中的第j个元素。

步骤2　应用熵值法计算准则层的指标权重。熵值法认为若某个指标下各目标属性值的分布较为接近，则该指标对于目标威胁评估的价值较低，其对应的指标权重较小；反之，若各目标属性值分布较为离散，则该指标对威胁评估的价值较高，其对应的指标权重也更大[15]。基于熵值法的指标权重计算过程如下：

首先，将根据3.1节计算得到的各目标威胁因素量化值进行归一化为

$$z_{ij}=\omega_{ij}\Big/\sum_{i=1}^{n}\omega_{ij},\ \ 1\le i\le n,\ \ 1\le j\le 5 \tag{5}$$

其中，ω_{ij}为第i个目标对于第j个威胁指标因素的量化值，z_{ij}为归一化的指标属性值。

其次，应用归一化的指标属性值信息，各指标的信息熵为

$$e_j=\frac{-1}{\ln n}\sum_{i=1}^{m}(z_{ij}\ln z_{ij}),\ \ 1\le j\le 5 \tag{6}$$

其中，e_j为第j个指标的信息熵。

最后，各评估指标在信息熵语义下的指标权重可以计算为

$$\omega_j^e=(1-e_j)\Big/\sum_{j=1}^{5}(1-e_j),\ \ 1\le j\le 5 \tag{7}$$

其中，ω_j^e为熵值法下第j个指标的权重。

步骤3　利用熵值法计算得到的指标权重对AHP准则层中指标权重进行修正

$$w_j=0.6\times w_j^{\mathrm{AHP}}+0.4\times w_j^e \tag{8}$$

其中，w_j为准则层中第j个指标的最终权重。

步骤4　计算方案层中各目标相对于准则层的指标权重。利用目标威胁因子向量构造方案层相对于准则层的重要性判别矩阵。令方案层各来袭目标导弹相对于准则层中第j个威胁评估指标的判别矩阵为$[b_{ik}^j]_{n\times n}$，该判别矩阵元素计算为

$$b_{ik}^j=\omega_{ij}/\omega_{kj}\ ,\ \ 1\le i\le n,\ \ 1\le k\le n \tag{9}$$

其中，b_{ik}^j为第j个指标下，第i个目标弹相较于第k个目标弹的重要程度。计算各指标下重要性判别阵的最大特征值对应的归一化特征向量$\boldsymbol{c}_j=[c_{1j},c_{2j},\cdots,c_{nj}]$，其中，$c_{ij}$为第$i$个目标弹在第$j$个指标下的威胁度。$\boldsymbol{c}_j$为方案层各目标相对于准则层中第$j$个指标的层次排序向量。

步骤5　计算目标的综合威胁度。结合准则层各指标修正后的权重和方案层各目标相对于准则层指标的层次排序结果，各目标的综合威胁度计算式为

$$v_i=\sum_{j=1}^{5}(c_{ij}\times w_j)\ ,\ \ 1\le i\le n \tag{10}$$

其中，v_i为第i个目标的综合威胁度。

4　基于DQN的WTA决策方法

基于DQN的WTA决策模型整体架构如图2所示，将火力分配过程看作一个多段决策过程，单步决策通过优化决策奖励值，实现对单个拦截弹的目标分配，通过依次对拦截弹进行分配决策，从而完成整个WTA过程。当完成一轮火力分配后，计算全局决策收益，并更新到临时记忆库中。DQN根据“均匀采样”策略利用临时记忆库中的分配经验(状态转移4元组)进行训练，不断完善Q网络，从而达到基于DQN的火力分配智能体可快速高效求解中大规模WTA问题的目的。

4.1　状态转移4元组设计

根据WTA问题特点，以火力单元数量的编号顺序作为决策时序，第i步决策表示对第i个拦截弹进行目标分配，即确定$x_{ij}=1$时j的取值。定义第i步决策的状态转移4元组为$<\boldsymbol{s}_i,\boldsymbol{a}_i,r_i,\boldsymbol{a}_{i+1}>$，其中$\boldsymbol{s}_i$为火力单元当前状态向量，包含第$i$步决策时的火力单元剩余量和当前火力单元的类型；$\boldsymbol{a}_i$为当前动作向量，表示将第$i$个拦截弹分配给指定的目标，包含第$i$步决策时选择的被分配目标编号和类型，及该目标已被分配的拦截弹数量；r_i为奖励函数，即采取相应动作所产生的奖励；$\boldsymbol{s}_{i+1}$为基于当前决策的下一步火力单元状态向量，即第i+1步决策时的火力单元剩余量和火力单元类型。

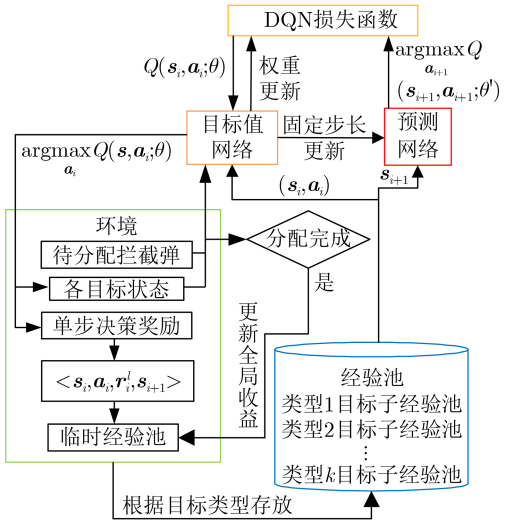

图 2 基于DQN的WTA决策模型

4.1.1 状态向量s_i定义

根据蓝方反导拦截系统的拦截弹资源配置和部署情况，构造合适的状态向量$\boldsymbol{s}_i$。由于不同类型的拦截火力单元对同一目标的毁伤概率存在差异，例如，美国的GBI和SM-3适用于拦截中高层目标，而PAC-3则擅长拦截低空大气层内的目标。因此，状态向量需包含火力单元的类型信息，同时也需要包含火力资源的剩余情况。由此，定义第i个火力单元分配时的状态为

$$\boldsymbol{s}_i=\left[1-\frac{m_{\text{cost}}}{m},\left[\frac{m_{\text{cost}}^1}{m},\frac{m_{\text{cost}}^2}{m},\cdots,\frac{m_{\text{cost}}^l}{m}\right],m_i\_\text{type}\right] \tag{11}$$

其中，m_{cost}为已分配的拦截弹数量；$m^i{}_{\text{cost}}$(i=1, 2,⋯,l)为第i类拦截弹已分配的数量；m_i_type为该拦截弹的类型独热编码。类似地，当执行完第i个火力单元分配后，更新状态信息，可得第i+1步决策时的状态量$\boldsymbol{s}_{i+1}$。值得注意的是，当i=m时，不存在$\boldsymbol{s}_{i+1}$。

4.1.2 动作向量a_i定义

在对拦截弹进行目标分配时，需要考虑目标的威胁度。目标威胁度越高，对应的打击优先级越高。当一个目标已被多个火力单元分配时，该目标的毁伤概率可以得到较好的保障，此时考虑给其分配火力单元的优先级随之下降。因此，在设计DQN的动作向量时，需要综合考虑目标威胁度、目标弹已被分配的情况。此外，由于同一拦截弹对不同类型的目标的毁伤概率各不相同，动作向量还需包括目标的类型信息。因此，假设第i步决策时，将拦截弹分配给第j个目标，可定义$\boldsymbol{a}_i$动作向量的一个决策动作$\boldsymbol{a}_{ij}$为

$$\boldsymbol{a}_{ij}=[[m_{\text{use}}^1,m_{\text{use}}^2,\cdots,m_{\text{use}}^l],\ v_j,\ n_j\_\text{type}] \tag{12}$$

其中，$m^i{}_{\text{use}}$ (i=1,2,⋯,l)为第i类火力单元分配到该目标上的数量；v_j为该目标的威胁度；n_j_type为该目标的类型独热编码。

4.1.3 奖励函数r_i定义

将最大化单步火力分配的毁伤概率增益作为单步决策奖励$\boldsymbol{r}_i^l$，定义为

$$\boldsymbol{r}_i^l=\Delta J_i=J^i-J^{i-1} \tag{13}$$

其中，J^i为第i步决策完成后的对敌方目标的整体毁伤概率，计算公式如式(1)所示。

若DQN只学习到上述单步决策奖励会导致DQN决策时出现“短视”现象，具体可描述为：在一轮火力分配的初期，DQN为了最大化单步决策奖励，会做出不利于最大毁伤概率的目标分配选择。假设有两个威胁度相同的目标，分别为目标1和目标2，拦截弹1和拦截弹2对目标1,2的毁伤概率分别为[0.86, 0.84]和[0.84, 0.75]。在基于DQN的WTA分段决策中，单步奖励最大化下的决策是将拦截弹1分配给目标1，拦截弹2分配给目标2，但按照最大化整体毁伤概率准则的分配结果是将拦截弹1分配给目标2，拦截弹2分配给目标1。造成这种冲突现象的原因在于DQN做当前决策时仅注重了单步决策奖励，忽视了全局收益，即并未考虑单步决策对后续拦截弹的分配决策带来的影响。由此，造成了本文所谓的“短视现象”。

为了改善上述这种现象，考虑单步决策对后续决策的影响，将代表一轮分配完成后的目标最终整体毁伤概率引入到单步决策的奖励函数中，兼顾火力分配的单步决策收益和全局收益，由此修正第i步决策的奖励函数为

$$\boldsymbol{r}_i=\alpha_i r_i^l+(1-\alpha_i)r_g\ ,\ \ 0<\alpha_i<1 \tag{14}$$

其中，α_i为权重系数，r_g=$J(x_{ij})$为目标整体毁伤概率。此外，将上式与仅考虑全局收益的奖励函数相比，可知由于引入了单步决策增益奖励，可以在一定程度上引导决策空间的探索，表现在能够使得搜索沿着在单步增益较大的空间内开展，提高搜索效率。因此，式(14)综合考虑单步和全局收益的奖励函数能够使得DQN兼顾优化解的全局性和搜索的快速性。

在火力分配初始阶段更容易发生“短视现象”，需要更加重视全局收益的影响，因此关于全局收益的权重系数需要设置的较大。当火力分配进入后期阶段时，由于大部分拦截弹已分配完成，最大化单步决策奖励下的决策逐步与最大化整体毁伤概率下的决策趋于一致，此时关于全局收益的权重系数可以适当减小，从而引导DQN进行快速探索。综上分析，本文采用动态权重的方法来实现上述目的，变权重系数的表达式为

$$\alpha_i=1-\mathrm{e}^{-3i/m} \tag{15}$$

4.2 “均匀采样”策略与经验存储

在完成所有火力单元的目标分配后，可通过火力分配决策矩阵按式(1)计算该轮火力分配的整体毁伤概率，并将其更新到该轮的各状态转移4元组中。由于不同类型的目标数量相差较大，导致对应各类目标的分配经验数量之间存在差异。若直接使用随机采样策略抽取样本进行训练，则会导致低数量类型的目标被抽取的概率较低，从而对该类目标的分配训练效果不佳。由此，本文采用根据目标类型进行抽取的“均匀采样”策略，将一轮火力分配完成后产生的分配经验按照目标类型分别进行存

储，训练时从各类型目标对应的子经验池中等量随机抽取一批经验，保证DQN能够等频率地学习到各类目标下的分配经验。

4.3 Q值迭代

对所有的m个拦截弹完成目标分配即完成了一轮的火力分配任务，因此定义本文火力分配场景中Q函数的最优贝尔曼方程为

$$Q^{\pi}(\boldsymbol{s},\boldsymbol{a})=\mathbb{E}_{\pi}\left[\sum_{k=i}^{m}\boldsymbol{r}_k|\boldsymbol{s}_i=\boldsymbol{s},\boldsymbol{a}_i=\boldsymbol{a}\right] \tag{16}$$

其中，$\boldsymbol{r}_k$为第k步分配决策的奖励。

由式(16)可得Q函数的更新规则为

$$\begin{aligned}Q(\boldsymbol{s}_i,\boldsymbol{a}_i)=&Q(\boldsymbol{s}_i,\boldsymbol{a}_i)+\alpha[\boldsymbol{r}_i\\&+\max_{\boldsymbol{a}_{i+1}}(\boldsymbol{s}_{i+1},\boldsymbol{a}_{i+1})-Q(\boldsymbol{s}_i,\boldsymbol{a}_i)]\end{aligned} \tag{17}$$

其中，α为学习率，$0<\alpha<1$。

为使DQN训练更加稳定，构造目标网络θ和预测网络θ'，两个网络的结构相同，初始权重相同[16]。利用式(18)和反向传播算法更新θ，θ'滞后若干决策步以后从θ复制节点权重进行更新

$$\text{Loss}=(\boldsymbol{r}_i+\max_{\boldsymbol{a}_{i+1}}Q(\boldsymbol{s}_{i+1},\boldsymbol{a}_{i+1};\theta')-Q(\boldsymbol{s}_i,\boldsymbol{a}_i;\theta))^2 \tag{18}$$

利用ε - greedy算法使DQN在决策空间探索和训练效率之间取得平衡。

综上，DQN训练流程主要包括：初始化训练配置参数；在ε - greedy机制下利用DQN模型选取最优拦截弹-目标对，并计算单步局部奖励r_i^l；一轮火力分配结束后计算目标群整体毁伤概率并根据式(14)更新该轮经验池的所有单步决策回报值；按照均匀采样策略等量抽取各目标类型的子经验池，进行目标网络训练，并按照预设间隔步数更新预测网络，对网络不断训练直至满足结束条件。

5 仿真测试与分析

5.1 目标威胁评估方法测试与分析

假定有10个来袭目标，其中目标1,2,5为近程弹道导弹，目标3,4,6为中程弹道导弹，目标7和8为远程弹道导弹，目标9和10为高超声速飞行器，各目标属性值如表1所示。

根据表1中数据，利用本文所提改进AHP方法计算评估指标权重，并与传统AHP方法的指标权重作对比，结果如表2所示。分析表1和表2结果可知，各目标弹的攻击地重要度指标分布较为分散，对拦截优先级判断的影响较大，因此，相较于传统的AHP方法，引入熵值法的改进AHP法对该指标因素给定的权重较大。相反，各目标弹的RCS值分布较为接近，对拦截优先级判断的影响较小，由改进AHP法计算得到的权重较小。因此，改进AHP方法可根据目标各威胁要素的量化指标分布情况，合理地调整指标权重，使得在威胁评估时突出不同目标间的差异性。

利用改进AHP法和传统AHP法对表1中各来袭目标弹进行综合威胁度计算，结果如表3所示，其中远程弹道导弹目标8的攻击地重要度最高，关机点速度大，因此两种方法都认为该目标的综合威胁度最高；而近程弹道导弹目标2的攻击地重要度和关机点速度最低，最大飞行高度低，因此两种方法计算该目标的综合威胁度都为最低。需要注意，相较于传统AHP方法，改进AHP方法认为高超声速目标9和10的目标威胁度更高，尤其是目标9的威胁度排序更加靠前。在实际战场中，高超声速目标通常杀伤力较大且难以拦截，威胁程度较高，改进AHP方法对高超声速飞行器的威胁评估结果更加符合实际。由此，可以说明本文提出的改进AHP

表 1 目标属性值

编号	攻击地重要度	剩余飞行时间(s)	最大高度(km)	关机点速度(km/s)	RCS (m^2)
1	4	220	260	2.3	0.007
2	9	250	225	2.1	0.005
3	4	530	630	4.2	0.012
4	2	550	680	4.8	0.013
5	6	240	235	2.2	0.010
6	2	610	710	5.1	0.015
7	1	1200	1600	6.8	0.017
8	0	1120	1450	6.6	0.016
9	2	1400	75	7.4	0.006
10	3	1500	78	7.1	0.007

表 2 传统和改进AHP方法的评估指标权重计算结果对比

	攻击地重要度	剩余飞行时间(s)	最大高度(km)	关机点速度(km/s)	RCS (m^2)
传统AHP	0.34	0.27	0.08	0.12	0.19
改进AHP	**0.44**	0.17	0.16	0.13	**0.10**

表 3 改进AHP与传统AHP法的目标威胁度评估结果

	编号				
	8	7	9	4	6
改进AHP法	0.125	0.119	0.111	0.107	0.106
传统AHP法	0.115	0.110	0.104	0.107	0.106
	编号				
	10	3	1	5	2
改进AHP法	0.104	0.095	0.091	0.078	0.060
传统AHP法	0.099	0.097	0.097	0.088	0.075

威胁评估方法的评价结果与实际情况更符合，具有较高的合理性。

5.2 DQN火力分配测试与分析

5.2.1 固定场景下的DQN火力分配测试与分析

针对表1中各来袭目标，利用本文所提DQN方法优化分配策略，DQN的训练参数设置为：学习率等于0.001，衰减率等于0.8，隐藏层数量为3，每层各100个节点，训练数据的批大小(batch_size)为32，预测网络的更新步长为50，共训练2000轮。设定拦截弹总量为20，低层、中层以及高层拦截弹的数量分别为11:6:3。其中，高层拦截弹对于远程目标的毁伤概率最大，为85%；对于中程目标的毁伤概率为40%。中层拦截弹对于中程目标的毁伤概率最大，为85%；而对于远程目标的毁伤概率为55%。低层拦截弹对于近程目标和高超声速目标具有较高的毁伤概率，分别为90%和55%。

经过2000轮的训练后，得到的DQN学习曲线如图3所示。由图3可知，在训练初期，由于ε值较小，DQN对决策空间进行随机探索，分配结果不稳定，随着训练回合数的增加，利用学习完善的DQN进行决策，整体毁伤概率逐渐上升并趋于稳定，最终稳定在0.91左右。火力分配结果如图4所示，该火力分配的整体毁伤概率为0.9128，由图3可知，对于威胁度最高的远程目标弹8，DQN分配了两枚针对性最强的高层拦截弹以及一枚近程拦截弹进行拦截，很大程度上确保毁伤该目标；对于威胁度较高的高超声速目标弹9，DQN则针对性地分配了3枚低层拦截弹，使该目标的毁伤概率达到90%以上；而对于威胁度最低的近程弹2,5,1，DQN则各分配了1枚低层拦截弹，既保证了目标的毁伤概率，也为拦截其他重要目标留出了较多的可支配火力资源。由此，说明DQN能够综合考虑目标威胁度、拦截弹-目标毁伤概率、火力资源配置情况，做出合理的火力分配决策。

此外，在上述场景下，对仅考虑全局收益的DQN火力分配模型进行训练，整体毁伤概率收敛曲线如图5所示。对比图3可知，当DQN仅考虑全局收益奖励时，DQN训练效率降低，收敛效果较差，从而使得最终的火力分配方案不佳。利用图5训练得到的DQN火力分配模型进行仿真测试，分配结果的整体毁伤概率较低，仅为0.678，火力分配结果不太理想。综上对比分析验证了式(14)综合考虑单步和全局收益的奖励函数设计可带来的训练效率和决策性能的提升。

5.2.2 随机场景下的DQN火力分配测试与分析

考虑实际作战场景中，目标规模通常难以准确预测，可用火力资源数量和配置也会随战场态势动态变化。因此，需要火力分配方法对WTA场景要素的变化具有较好的鲁棒性。考虑目标-拦截弹数量变化，毁伤概率和目标威胁度在小范围内浮动的WTA随机场景下，对DQN火力分配模型进行训练。每一轮的训练场景中，目标数量和拦截弹数量分别为[20, 30]和[30, 60]之间的随机整数，其中近、中和远程目标数量分别占目标总量的20%～40%，20%～40%，10%～20%，其余为高超声速目标。

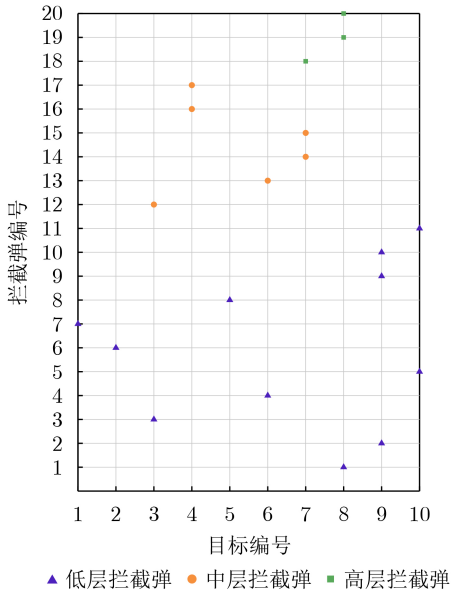

图 4 固定场景下DQN火力分配方案

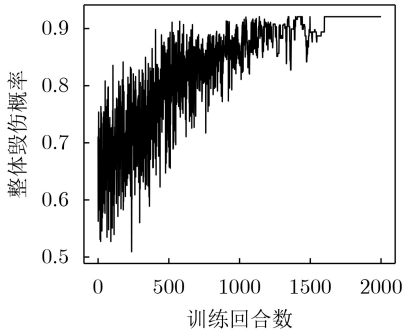

图 3 固定场景下DQN训练效果

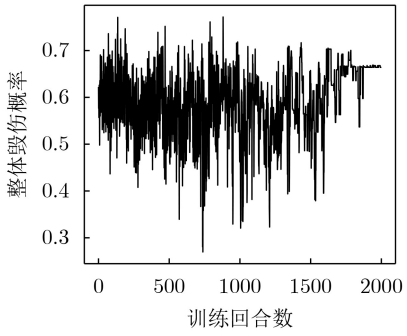

图 5 固定场景下仅考虑全局收益的DQN训练效果

低、中层拦截弹配比范围均为30%～40%，剩余为高层拦截弹。

为体现DQN在随机场景下的训练效果，对训练过程进行1000次蒙特卡罗仿真，得到的DQN平均学习收敛曲线如图6所示。从图中可以看出，DQN能够在场景要素变化的情况下进行有效训练，随着训练的进行，平均整体毁伤概率逐步提高并最终收敛于0.9左右。该结果可以说明本文所提DQN方法在WTA要素变化的场景下具备良好且稳定的训练效果。

为了验证本文所提基于DQN的火力分配算法的性能，利用上述训练得到的DQN火力分配模型与文献[10]中的基于改进粒子群算法(Particle Swarm Optimization, PSO)的WTA优化方法，以及基于目标威胁度的随机分配法进行比较。其中，PSO方法的种群规模设为60，迭代次数为5000；随机法可描述为针对第j个可用火力单元，产生[0,1]之间的随机数x_{rand}^{j}，若满足

$$x_{\text{rand}}^{j}\in\Big(\sum_{1}^{i}\omega_i,\ \sum_{1}^{i+1}\omega_i\Big],\ i=0,1,\cdots,m,\ \omega_0=0 \tag{19}$$

则将该火力单元分配给第i+1个目标，其中ω_i为归一化的目标威胁度。该分配方法使得火力单元有更大的概率分配给威胁度较高的目标。

设置如表4所示的3个测试用例，测试时的毁伤概率各类型目标数量占比和各类型拦截弹数量占比的设定与训练场景保持一致。在训练场景中，目标数量和拦截弹数量分别在[20,30]和[30,60]之间随机取值，测试用例1是一个较小规模的WTA场景，目标和拦截弹数量分别为15和25，目标和拦截弹的数量规模均低于DQN模型训练时的各自最小规模；用例3是一个较大规模的WTA场景，目标、拦截弹数量分别为35和50，其目标数量规模大于DQN模型训练时的最大规模。

3种方法在不同测试场景下产生的目标群整体毁伤概率和运行时间如表5所示，随着WTA规模的增大，基于改进PSO方法的搜索空间规模爆炸式增长，受限于种群规模和迭代次数，所得解的质量不断下降，尤其在用例3中，由于搜索空间的急剧增大，该方法求解得到的整体毁伤概率下降到了0.75左右，且耗时很长，难以满足高动态场景下火力分配决策的快速性需求。而基于DQN的火力分配模型得益于充分的训练，基于良好的网络参数，能够适应目标和火力资源配置动态变化的情况，在3个测试用例下都能保持较好的求解质量，尤其是在用例3，较大规模的火力分配问题中也能保持0.85以上的毁伤概率，且能够满足决策快速性需求。此外，用例1和用例3的测试结果表明，模型能够适应超出训练场景参数范围的WTA场景，因此，基于训练得到的DQN模型对于非预期内的场景参数变化情况，包括目标和拦截弹数量、毁伤概率等变化情况，具有一定的鲁棒性，可适用于战场中的突发动态情况下的火力分配应用。

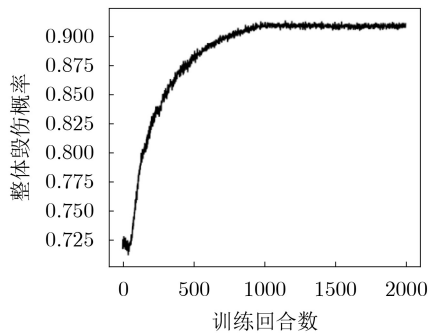

图 6 1000次蒙特卡罗仿真训练

表 4 测试用例参数

测试用例编号	目标数量比	拦截弹数量比
#1	5:5:3:2	12:8:5
#2	10:8:5:2	18:15:12
#3	12:9:9:5	25:15:10

表 5 3种场景测试结果

指标	测试用例编号	分配方案求解方法		
		DQN	PSO	随机法
整体毁伤概率	#1	0.921	0.982	0.620
	#2	0.918	0.907	0.590
	#3	**0.856**	**0.758**	0.540
运行时间(s)	#1	0.050	22.001	0.001
	#2	0.170	62.021	0.003
	#3	**0.220**	**137.000**	0.019

6 结束语

本文考虑由不同性能拦截弹组成的一体化导弹防御系统对不同类型的来袭目标群实施火力分配的问题，提出了一种融合改进AHP和DQN的WTA优化方法。首先，应用基于熵值法的改进AHP方法评估来袭目标威胁度，本文方法由于引入了目标威胁指标量化数据的分布差异，相较于典型的AHP方法能够较好地突出区分目标威胁差异，结果具有良好的合理性。接着，针对基于传统启发式方法求解中大规模WTA问题效率低、优化解质量不高的问题，本文在DQN框架下将WTA过程看作一个多段决策过程，通过设置可综合兼顾训练效率和决策性能的奖励函数，引入公平采样策略等手段，建立了基于DQN的火力分配方法。大量仿真结果表

明，在固定和随机的WTA场景下，本文提出的基于DQN的WTA优化方法均能在较少的训练次数下快速收敛，针对不同的测试用例均能给出较优的火力分配方案，且对于WTA场景参数动态变化具有一定的适应性，具备对战场环境动态变化的适应能力。同时，相较于经典的PSO算法，本文算法在处理中大规模WTA问题时优势明显，具备决策的快速性和准确性。

参 考 文 献

[1] KLINE A, AHNER D, and HILL R. The weapon-target assignment problem[J]. *Computers & Operations Research*, 2019, 105: 226–236. doi: 10.1016/j.cor.2018.10.015.

[2] YUE Jiao and ZHANG Ke. Vulnerability Threat assessment based on AHP and fuzzy comprehensive evaluation[C]. 2014 IEEE Seventh International Symposium on Computational Intelligence and Design, Hangzhou, China, 2014: 513–516. doi: 10.1109/ISCID.2014.231.

[3] 杨罗章, 胡生亮, 冯士民. 基于Entropy-TOPSIS方法的目标威胁动态评估与仿真[J]. 兵工自动化, 2020, 39(3): 53–56,60. doi: 10.7690/bgzdh.2020.03.012.
YANG Luozhang, HU Shengliang, and FENG Shimin. Dynamic evaluation and simulation of targets threat based on entropy and TOPSIS method[J]. *Ordnance Industry Automation*, 2020, 39(3): 53–56,60. doi: 10.7690/bgzdh.2020.03.012.

[4] 陈龙, 马亚平. 基于分层贝叶斯网络的航母编队对潜威胁评估[J]. 系统仿真学报, 2017, 29(9): 2206–2212,2220. doi: 10.16182/j.issn1004731x.joss.201709044.
CHEN Long and MA Yaping. Threat assessment of aircraft carrier formation based on hierarchical Bayesian network[J]. *Journal of System Simulation*, 2017, 29(9): 2206–2212,2220. doi: 10.16182/j.issn1004731x.joss.201709044.

[5] 杨爱武, 李战武, 徐安, 等. 基于RS-CRITIC的空战目标威胁评估[J]. 北京航空航天大学学报, 2020, 46(12): 2357–2365. doi: 10.13700/j.bh.1001-5965.2019.0638.
YANG Aiwu, LI Zhanwu, XU An, *et al*. Threat assessment of air combat target based on RS-CRITIC[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2020, 46(12): 2357–2365. doi: 10.13700/j.bh.1001-5965.2019.0638.

[6] LLOYD S P and WITSENHAUSE H S. Weapon allocation is NP-Complete[C]. The IEEE Summer Simulation Conference, Reno, USA, 1986: 1054–1058.

[7] 王邑, 孙金标, 肖明清, 等. 基于类型2区间模糊K近邻分类器的动态武器目标分配方法研究[J]. 系统工程与电子技术, 2016, 38(6): 1314–1319. doi: 10.3969/j.issn.1001-506X.2016.06.15.
WANG Yi, SUN Jinbiao, XIAO Mingqing, *et al*. Research of dynamic weapon-target assignment problem based on type-2 interval fuzzy K-nearest neighbors classifier[J]. *Systems Engineering and Electronics*, 2016, 38(6): 1314–1319. doi: 10.3969/j.issn.1001-506X.2016.06.15.

[8] 王净, 战凯, 晏峰. 基于动态规划算法的舰空导弹火力分配模型研究[J]. 舰船电子工程, 2011, 31(2): 24–26. doi: 10.3969/j.issn.1627-9730.2011.02.007.
WANG Jing, ZHAN Kai, and YAN Feng. Ship-to-air missile firepower-distributing model study based on dynamic programming algorithm[J]. *Ship Electronic Engineering*, 2011, 31(2): 24–26. doi: 10.3969/j.issn.1627-9730.2011.02.007.

[9] 丁立超, 黄枫, 潘伟. 基于改进混沌遗传算法的炮兵火力分配方法[J]. 系统仿真技术, 2021, 17(1): 12–16. doi: 10.16812/j.cnki.cn31-1945.2021.01.003.
DING Lichao, HUANG Feng, and PAN Wei. Artillery fire allocation method based on improved chaotic genetic algorithm[J]. *System Simulation Technology*, 2021, 17(1): 12–16. doi: 10.16812/j.cnki.cn31-1945.2021.01.003.

[10] 李俨, 董玉娜. 基于SA-DPSO混合优化算法的协同空战火力分配[J]. 航空学报, 2010, 31(3): 626–631.
LI Yan and DONG Yu'na. Weapon-target assignment based on simulated annealing and discrete particle swarm optimization in cooperative air combat[J]. *Acta Aeronautica et Astronautica Sinica*, 2010, 31(3): 626–631.

[11] SILVER D, SCHRITTWIESER J, SIMONYAN K, *et al*. Mastering the game of Go without human knowledge[J]. *Nature*, 2017, 550(7676): 354–359. doi: 10.1038/nature24270.

[12] ZHU Yuke, MOTTAGHI R, KOLVE E, *et al*. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]. 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017: 3357–3364. doi: 10.1109/ICRA.2017.7989381.

[13] 施伟, 冯旸赫, 程光权, 等. 基于深度强化学习的多机协同空战方法研究[J]. 自动化学报, 2021, 47(7): 1610–1623. doi: 10.16383/j.aas.c201059.
SHI Wei, FENG Yanghe, CHENG Guangquan, *et al*. Research on multi-aircraft cooperative air combat method based on deep reinforcement learning[J]. *Acta Automatica Sinica*, 2021, 47(7): 1610–1623. doi: 10.16383/j.aas.c201059.

[14] 阎栋, 苏航, 朱军. 基于DQN的反舰导弹火力分配方法研究[J]. 导航定位与授时, 2019, 6(5): 18–24. doi: 10.19306/j.cnki.2095-8110.2019.05.003.
YAN Dong, SU Hang, and ZHU Jun. Research on fire distribution method of anti-ship missile based on DQN[J]. *Navigation Positioning and Timing*, 2019, 6(5): 18–24. doi: 10.19306/j.cnki.2095-8110.2019.05.003.

[15] ZHU Yuxin, TIAN Dazuo, and YAN Feng. Effectiveness of entropy weight method in decision-making[J]. *Mathematical Problems in Engineering*, 2020, 2020: 3564835. doi: 10.1155/2020/3564835.

[16] MNIH V, KAVUKCUOGLU K, SILVER D, *et al*. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533. doi: 10.1038/nature14236.

谢俊伟：男，博士生，研究方向为智能决策与控制.

方　峰：男，讲师，博士，研究方向为飞行器协同制导与控制、智能决策.

彭冬亮：男，教授，博士，博士生导师，研究方向为信息融合、检测与估计.

任金磊：男，工程师，硕士，研究方向为飞行器设计、弹道导航制导控制、智能控制.

王昌平：男，硕士生，研究方向为导弹协同制导.

责任编辑：余　蓉