doi:10.11887/j.cn.201402027 http://journal.nudt.edu.cn

# 基于马尔科夫决策的目标选择策略*

雷 霆[1,2],朱 承[1],张维明[1]
(1.国防科技大学 信息系统工程重点实验室,湖南 长沙 410073;2.军事科学院 运筹所,北京 100091)

**摘 要**:目标选择是军事计划的关键要素之一。基于马尔科夫决策方法,解决具有复杂目标间关联的多阶段目标选择问题。使用与或树描述目标体系各层状态间的影响关联,并以目标体系整体失效为求解目的,建立了基于离散时间MDP的多阶段打击目标选择模型。在LRTDP算法基础上提出一种启发式方法,通过判断从当前目标体系状态到达体系失效状态的演化过程中的可能资源消耗和失败概率,来提供对当前状态的评估值,该方法能有效排除问题搜索空间中不能到达体系失效目的的中间状态,压缩了由于目标间复杂关联而增长的巨大状态空间。用实验验证了该方法有效性,实验结果表明,该方法直观实用,对目标间具有复杂关联关系的目标打击决策有一定参考价值。

**关键词**:目标选择;目标体系;与或树;离散时间马尔科夫决策过程
**中图分类号**:E917 **文献标志码**:A **文章编号**:1001-2486(2014)02-0161-07

# Research on the method of target selecting policy based on the Markov decision process

*LEI Ting*[1,2] *ZHU Cheng*[1], *ZHANG Weiming*
(1. Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China;
2. Institute of Military Operation Research, Academy of the Military Science, Beijing 100091, China)

**Abstract**: Target selecting is an important aspect of military operational planning. The Markov Decision Process(MDP) method was used to solve the multi-phase target selecting problem which has complex relations among targets. Firstly, the and-or tree was used to describe the relations among the layers of the target system of system(TSoS), and a Discrete Time Markov Decision Process(DTMDP) method was proposed for modeling target selecting whose objective was to neutralize the TSoS. Secondly, an LRTDP algorithm based heuristic was proposed to give the estimate value of the current state of the TSoS, which was calculated by considering the potential resource consumption and failure probability of the evolution process from the current state to the lapse state of the TSoS, and the heuristic can effectively exclude the intermediate states which cannot be transferred to the lapse state, in order to reduce the huge search space of the model because of the complex relations among targets. Finally, a case was proposed to validate the method. The results show that the method is intuitive and practical, and can facilitate the target selecting decision making when there are complex relations among the targets.

**Key words**: target selecting; target system of system; and-or tree; discrete time Markov decision process.

目标选择是军事决策过程的重要组成部分,现代战争中的目标选择问题要置于打击目标体系的作战过程中分析。

目标体系(Target System of System,TSoS)是由多个作战系统构成的集合,每个作战系统实现一定任务并对体系使命产生影响[1]。打击目标体系的目的是使体系崩溃,打击过程由于存在资源约束等原因被划分为多个阶段,因此如何打击目标体系是具有复杂目标关联的多阶段目标选择问题。

传统目标选择方法多是通过层次分析法[2]、多属性决策理论[3]等对目标进行评估和排序,没有考虑目标间复杂关联,为处理该问题,目前主要采用贝叶斯网络[4]、影响网[5]、影响图[6]和图论[7]、故障树方法[8-9]描述目标体系内影响关联。但以上方法均未考虑目标选择的多阶段决策特征,没有利用行动中间结果调整目标。

目标选择的动态性在动态武器目标分配问题和军事行动规划问题中得到研究。蔡怀平等[10]研究了动态武器目标分配问题中的马尔科夫性,

* **收稿日期**:2013-07-16
**基金项目**:国家自然科学基金资助项目(61273322,71001105,91024006)
**作者简介**:雷霆(1981—),男,安徽宿州人,博士研究生,E-mail:leiting_nudt@126.com;
张维明(通信作者),男,教授,博士生导师,E-mail:zhangweiming@nudt.edu.cn

解武杰等[11]将马尔可夫过程用于分析防空武器目标选择策略;Boutilier 等在马尔科夫决策过程(Markov Decision Process,MDP)基础上提出决策理论规划方法[12],对具有阶段决策的军事行动进行建模[13-14],但没有考虑目标关联和相应的复杂打击效果,不能直接用于求解打击目标体系过程中的目标选择问题。阳东升等[15]利用动态贝叶斯网络描述了战场重心及作战行动间影响关系,但搜索空间很大时求解效率不高,王长春等[16]利用复杂网络仿真方法分析体系对抗过程,但是建模过程较复杂。

## 1 目标选择问题描述

为分析目标选择问题,需分析打击目标对目标体系状态的影响。与或树使用图形化能将复杂问题分解为多个简单子问题,因此使用与或树描述体系中状态间的影响关系。

目标体系的状态包括三类要素状态:目标单元状态 $G^T$、目标系统能力状态 $G^N$ 和目标体系能力状态 $G^S$。

目标单元是目标体系中最基础的要素,能被直接摧毁,如单部雷达,其状态用叶节点集 $G^T=\{g_i^T\}(1\leqslant i\leqslant I)$ 描述,$I$ 为目标单元数量,单元毁伤,$g_i^T=1$;单元正常,$g_i^T=0$。

目标系统是多个目标单元或子系统的集合,之间相互关联,显现某种作战能力,如预警能力。其状态用非终端节点集 $G^N=\{g_j^N\}(1\leqslant j\leqslant J)$ 描述,$J$ 为目标系统数量,系统能完成任务,$g_j^N=1$;不能完成任务,$g_j^N=0$。其包含的目标单元和子系统能力状态作为其在与或树中子节点,通过逻辑与、或关系,对系统能力状态产生影响。

目标体系是多个目标系统的集合,体现出支持某个使命的能力,如防空使命能力。体系能力状态使用根节点 $G^S$ 描述,体系能达成使命,$G^S=1$;不能达成,$G^S=0$。其包含的各目标系统能力作为其子节点,通过逻辑与、或关系对体系能力状态产生影响。

在与或树中,设节点 $g\in G$,其子节点集 $SG=\{sg_k\}(1\leqslant k\leqslant K)$($K$ 为子节点数量),满足下式(1)时,$SG$ 为 $g$ 的与子节点集。

$$g=\bigcap_{k=1}^{K}sg_k=\prod_{k=1}^{K}sg_k \tag{1}$$

满足下式(2)时,则称 $SG$ 为 $g$ 的或子节点集。

$$g=\bigcup_{k=1}^{K}sg_k=1-\prod_{k=1}^{K}(1-sg_k) \tag{2}$$

由目标单元状态对应叶节点 $G^T=\{g_i^T\}$ 的取值,根据式(1)(2),能得到与或树所有节点取值,设函数关系为 $\Phi$:

$$G=\Phi(g_1^T,\cdots,g_i^T,\cdots,g_I^T) \tag{3}$$

目标体系状态影响关系的与或树如图 1 所示,子节点间绘制弧线表示与关系,未绘制表示或关系。

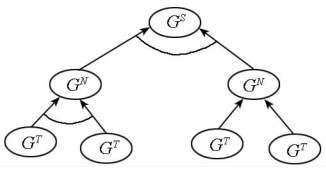


图 1 目标体系状态影响关系图

Fig. 1 The relationships among the states of TSoS

目标选择问题可描述为:已知目标体系状态影响关系和打击阶段、资源约束,打击目标的资源消耗、成功概率,求解在每个阶段的目标选择策略,使得达成体系失效的可能性最大。

## 2 目标选择过程建模

### 2.1 问题假设

(1)打击目标体系过程分为若干个作战阶段,使用有限资源,目的是使体系失效;

(2)目标体系状态为进攻方完全感知,目标选择决策仅与当前阶段状态有关,在当前状态被观察后,进攻方选择打击目标;

(3)打击每个目标具有一定成功概率,消耗一定资源,每个阶段打击多个目标,使得目标体系状态在下一阶段发生概率迁移。

### 2.2 目标选择决策模型

在符合以上假设时,打击过程中目标体系状态的变化可认为是一个离散时间随机过程,其变化过程的状态转移概率由打击目标行动所控制,因此目标选择决策成为一个离散时间马尔科夫决策过程,其最优决策就是每阶段要选择打击哪些目标,使目标体系失效的概率最大化。

本文使用 DTMDP 模型描述打击目标体系的目标选择决策过程,即以下多元组:

$$<S,S_0,S_T,A,P,C>$$

$S$ 是有限状态集,$S=\{(t,R,G)\}$,$t$ 指当前第 $t$ 阶段,$R=(R_1,\cdots,R_k,\cdots,R_K)$ 描述资源的状态向量,$R_k$ 为第 $k$ 类资源数量,$G=(g_1^T,\cdots,g_I^T,g_1^N,\cdots,g_J^N,G^S)$,表示体系的状态向量。

$S_0$ 是初始状态。$S_T$ 是终止状态集,对应于资源、时间消耗完毕,或目标体系失效的状态,在此状态下打击过程结束。

$A$ 是所有行动组成的有限集,$A(s)$是在状态 $s$ 下可采取的行动集,$a \in A(s)$包含多个目标单元打击任务 $\{Task_i\}$ $(1 \leqslant i \leqslant I)$,$Task_i$ 成功概率为 $P_i$,即 $P_i(G_i^{\mathrm{T}}=1 \mid Task_i)=P_i$。若 $R_k(s,Task_i)$表示 $Task_i$ 在状态 $s$ 下消耗第 $k$ 种资源的数量,$L_k$ 表示第 $k$ 种资源在每阶段的最大允许使用数量,则:

$$\sum_{i=1}^{I} R_k(s,Task_i) < \min(L_k,R_k)$$

$P: S \times S \times A \rightarrow [0,1]$是在可用行动 $a$ 下状态转移 $s \rightarrow s'$的概率函数,表示在打击行动 $a$ 下,状态在下一阶段变化的可能性。

在行动 $a$ 作用下,状态变化为 $s'=(t',R',G')$,包括:

(1)阶段状态变化

$$t=t+1 \tag{4}$$

(2)资源状态的变化

$$R' = R - \sum_i R(s,Task_i) \tag{5}$$

(3)目标体系要素的状态变化

目标体系状态变迁函数为各要素状态变迁概率函数的连乘积,根据目标体系层次间影响关系和式(3),其变迁函数为:

$$\begin{aligned} P(G' \mid G,a) &= P(\Phi(G^{\mathrm{T}'}) \mid a) \\ &= \prod_i P_i(\Phi(G^{\mathrm{T}'}_i) \mid Task_i) \end{aligned} \tag{6}$$

求解状态变迁概率的计算流程如图 2:

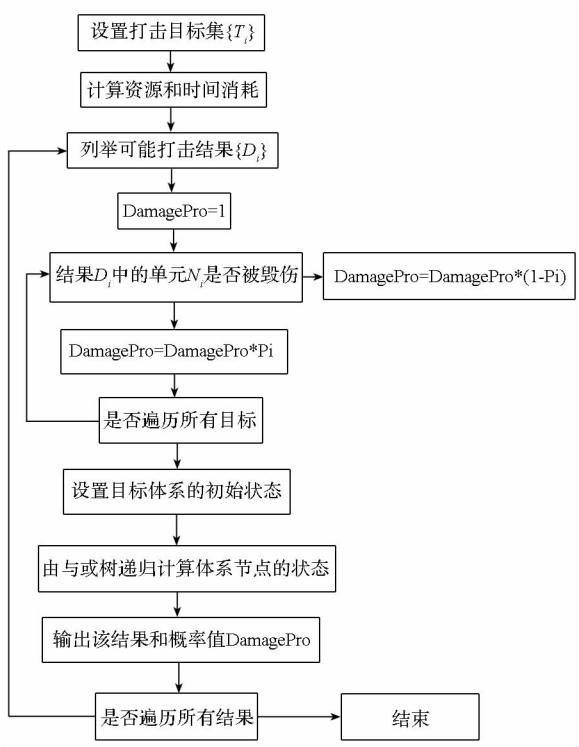


图 2　求解状态变迁概率的流程

Fig. 2　Calculate process of change probability of state

$C: S \times A \rightarrow C$ 是费用函数,$C(s,a)$是状态 $s$ 时采取行动 $a$ 所花费费用,即一步转移费用。

打击行动费用包括打击目标消耗资源 $RC$、时间 $TC$ 和到达目的状态(目标体系失效)的失败概率 $FC$[10]。在此给出下式:

$$\begin{aligned} C(s,a) = {} & w_1 \cdot FC(s,a) \\ & + w_2 \cdot RC(s,a) \\ & + w_3 \cdot TC(s,a) \end{aligned} \tag{7}$$

式中每个代价组成都有着对应权重值,其取值范围为:

$$\begin{cases} w_i \in [0,1] \quad i=1,2,3 \\ \sum_{i=1}^{3} w_i = 1 \end{cases} \tag{8}$$

权重值的选取和具体研究问题特点相关,针对具体目标体系,不同用户所关注的指标重要性不同,由用户调整权重的分配。

所要求取的是一个目标选择策略 $\pi(s): S \backslash S_T \rightarrow A$,策略将体系打击过程的状态映射到行动决策,即在每个打击阶段如何根据状态选择要打击的目标,使得最终能够达成目标体系失效可能性最大,同时所花费代价总和为最小。

### 2.3　模型复杂度分析

打击目标体系过程中的目标选择模型和以往基于 MDP 的目标选择或军事计划模型[12-14]存在着以下区别:

(1)问题假设不同。以往模型中假设目标间无关联,而本模型假设目标间相互影响;(2)终止状态不同。以往模型是以最大化毁伤目标为期望值,而本模型是以达成目标体系失效为目的;(3)状态空间不同。以往模型的状态空间是所有目标的状态,而本模型的状态空间包含了目标单元、系统能力、体系能力三类要素状态,使得状态空间复杂度增加;(4)时间尺度不同。以行动阶段而非具体时间来描述打击目标体系过程,并假设行动能够在单阶段内完成,简化了行动空间描述;(5)状态迁移函数不同。以往模型只需计算各目标的状态迁移,而本模型中的状态迁移还需考虑不同层次间要素的状态影响关系。

设该模型状态数量上界为 $N_f$,$T$ 是最大阶段数,目标单元数量为 $I$,$K$ 是资源类型数量,$R_k$ 是第 $k$ 类资源数量,则有:

$$N_f < T2^I \prod_{k=1}^{K} R_k \tag{9}$$

由于目标系统、体系能力状态与目标单元状态存在映射关系,因此目标体系的状态数等于目标单元的状态数 $2^I$。

例如目标体系有 10 个目标单元,最多可用 10 个打击阶段,打击目标行动消耗 2 类资源 $A$ 和 $B$,数量分别为 15、20 个单位,则模型中的状态数上限为:

$$N_f < 10 \cdot 2^{10} \cdot 15 \cdot 20 = 3\ 072\ 000$$

## 3 求解算法

### 3.1 求解框架

本问题状态空间巨大,并且只关注求解从目标体系初始状态到达终止状态的行动策略,而 MDP 值迭代或策略迭代方法需对全状态空间进行遍历,因此求解效率较低,这就需要使用启发式搜索算法来求解。RTDP(Real Time Dynamic Programming)[18] 的改进算法 LRTDP(Labeled RTDP)方法要比其他如 LAO * 等求解 MDP 的启发式搜索算法要更有效率[19],因此本文使用 LRTDP 方法求解该模型。

RTDP 是基于试验(trials-based)的方法,每次试验从初始状态开始,基于当前状态值的启发式,根据贪婪策略选择行动,然后根据行动的概率结果随机创建后续状态,直至到达目的状态,然后进行反向值迭代。

所用值迭代公式为[18]:

$$Q_{t+1}(s,a) = c(s) + \sum_{s' \in S} \Pr[s' \mid s,a] V_t(s) \tag{10}$$

当 $V_t(s_0)$ 到 $V_{t+1}(s_0)$ 没有变化时,可认为 $Q(s,a)$ 收敛。

LRTDP[19] 在反向值迭代过程中,会标记所经历的状态是否被求解,避免了重复计算已求解过的状态,因此加快了收敛。

### 3.2 启发式

文献[13]中设计了基于行动成功概率、行动执行时间和资源边界的启发式提供对 $V_0(S)$ 的最佳估计值,使得对所有状态 $s$,$V_0(S) < V(S)$,以促进 LRTDP 中算法的收敛,但由于打击目标体系过程中的目标选择模型和传统规划模型在状态空间、迁移函数上的区别,该启发式不能直接应用于前者。

针对打击目标体系过程特点,分别设计新的启发式来计算从目标体系当前状态 $S$ 到达目标体系失效状态的最小失败概率 $\min V(S,fail)$ 和最小资源消耗需求 $\min V(S,resource)$,并进行加权组合,以得到对 $V_0(S)$ 的最佳估计值。

和文献[13]中启发式考虑了时间代价不同,由于打击目标的时间消耗为单个阶段,从当前状态到达目标体系失效状态的最小时间消耗需求 $\min V(S,time)$ 总是为单个阶段,因此在新启发式中没有考虑时间代价。

(1)到达目标体系能力失效状态的最小失败概率

为判断从当前状态到达体系失效状态的最小失败概率,先求得最大成功概率,即从当前状态下预期能采取的所有打击目标行动能够达成的体系失效概率。

当目标体系与或树中非叶子节点 $g$ 具有子节点集 $SG = \{sg_k\}$ $(1 \leqslant k \leqslant K)$($K$ 为子节点数量)时,当 $SG$ 为与关系时,使 $g$ 失效的最大成功概率 $Pro$ 为:

$$Pro = \prod_{k=1}^{K} (sg_k \cdot Pro_k) \tag{11}$$

当 $SG$ 为或关系时:

$$Pro = 1 - \prod_{k=1}^{K} [(1 - sg_k)(1 - Pro_k)], 1 \leqslant k \leqslant K \tag{12}$$

其中 $Pro_k$ 表示使得第 $k$ 个子节点失效的最大成功概率,$sg_k$ 描述第 $k$ 个子节点是否失效,失效时取 1,正常时取 0。

其基本过程为:

1)与或树自根节点向下遍历各节点;

2)取得各节点的状态,当节点状态为失效,则该节点的毁伤概率为 1,当节点状态为正常,取得其所有子节点的失效概率值,根据子节点间的与或关系计算使该节点失效的概率值;

3)直至遍历至叶节点,获得对应打击目标行动的成功概率(即节点失效概率值),然后递归计算使根节点失效的成功概率值。

用 1 减去使根节点失效的最大成功概率值即得到使目标体系失效的最小失败概率。

(2)到达目标体系失效状态的最小消耗

为求解到达目标体系失效状态的最小消耗资源,我们假设从当前状态开始,所采取的每次打击行动都能成功摧毁目标。根据与或树的结构层次计算能够导致目标体系失效所需的行动集的最小消耗资源。

当目标体系与或树中非叶子节点 $g$ 具有子节点集 $SG = \{sg_k\}$ $(1 \leqslant k \leqslant K)$($K$ 为子节点数量)时,当 $SG$ 为与关系时,使 $g$ 失效的最小资源消耗 $Res$ 为:

$$Res = \sum_{k=1}^{K} (1 - sg_k) \cdot Res_k \tag{13}$$

当 $SG$ 为或关系时:

$$Res = \min(\{(1 - sg_k) \cdot Res_k\}), 1 \leqslant k \leqslant K \quad (14)$$

其中 $Res_k$ 表示使得第 $k$ 个子节点失效的最小资源消耗,$sg_i$ 描述第 $k$ 个子节点是否失效,失效时取 1,正常时取 0。

其基本过程为:

1)与或树自根节点向下遍历各节点;

2)当节点状态为失效,则该节点资源消耗为 0,当节点状态为正常,则取得其所有子节点消耗资源值,根据子节点间与或关系综合得到该节点资源消耗值;

3)直至遍历到叶节点,获得对应打击目标行动的消耗资源,然后递归计算使根节点(体系能力)失效的资源消耗值。

## 4　算例

本文实验程序在 Aberdeen 等[13]的 LRTDP 算法的基础上,考虑了目标状态之间的相互影响关系和行动造成的复杂效果,加入与或树推理程序来计算行动产生的体系状态和其对应的概率,并以体系整体状态而非具体目标毁伤作为判断打击效果标准。

在实验中对文献[13]中无启发式的算法和本文的启发式算法两种情况进行对比。为在结果策略中体现用户对不同评估指标的偏好,还分析了代价权重调整对结果策略评价指标的影响。

实验硬件环境是 Intel Core Duo T8100 2.1GHz CPU,内存是 2GB,程序编译环境为 CodeBlocks10.05。

该案例如图 3 所示:体系 SoS 包含指挥系统 C 和防空系统 M,指挥系统 C 包含主指挥所系统 C1、备份指挥所 C2,主指挥所系统 C1 包含指挥通信设施 F、电力系统 E,电力系统包括输电设施 T、发电站系统 G1、发电站 G2,发电站系统 G1 包括油气管道 P、发电机组 K,防空系统 M 包括防空阵地系统 M1、防空阵地 M2,防空阵地系统 M1 包括防空导弹 D、防空指挥系统 S,防空指挥系统包括防空雷达 R、指挥车 B。具备与关系的节点之间用弧线进行连接,没有用弧线连接的节点为或关系。

假设打击目标使用两类资源 A、B,资源 A 全局数量约束为 38,阶段约束为 10;资源 B 全局约束为 36,阶段约束为 8。打击目标单元所消耗资源如表 1 所示。

**表 1　打击目标单元的行动参数**

Tab. 1　Parameter of the target unit striking action

| 目标 | 毁伤率 | 资源 A | 资源 B |
|---|---|---|---|
| D | 0.8 | 6 | 4 |
| M2 | 0.9 | 7 | 5 |
| F | 0.75 | 6 | 4 |
| C2 | 0.6 | 6 | 6 |
| K | 0.85 | 5 | 5 |
| G2 | 0.7 | 5 | 5 |
| B | 0.8 | 6 | 4 |
| T | 0.75 | 6 | 5 |
| R | 0.9 | 6 | 4 |
| P | 0.8 | 5 | 4 |

根据有无启发式和权重设置,分为 17 组算例进行实验,权重设置顺序为资源代价、时间代价、失败代价。每种情况下求得策略都运行 10000 次仿真进行评估,计算每个策略值的指标平均值如表 2 所示。

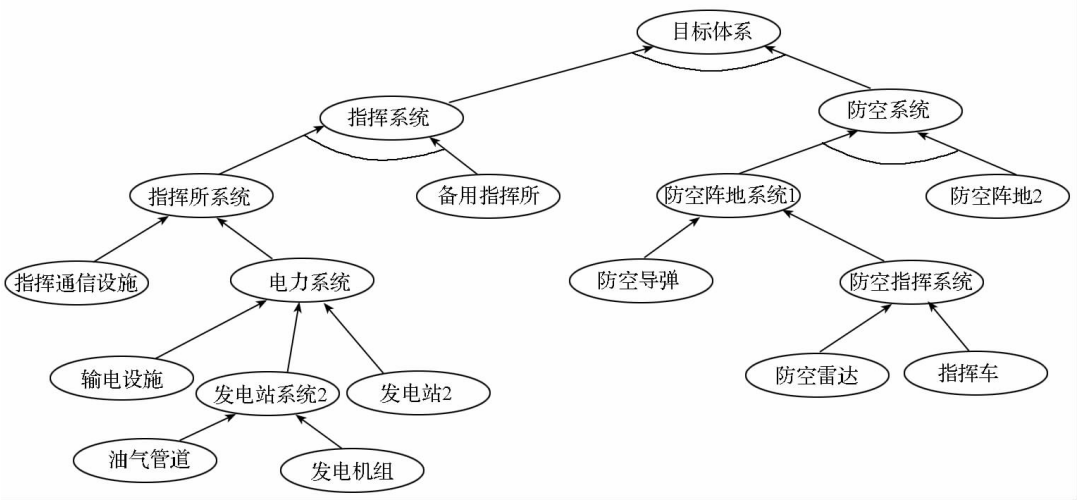


图 3　目标体系结构

Fig. 3　Structure of the TSoS

表 2 不同算例的计算结果

Tab. 2 Result of different conditions

| 序号 | 权重设置 | 启发式 | 资源消耗 | 打击阶段 | 失败概率 | 访问状态 | 算法时间(s) |
|---|---|---|---|---|---|---|---|
| 1 | (0.333,0.333,0.334) | 无 | 52.7534 | 4.9371 | 0.30054 | 182600 | 94 |
| 2 | (0.333,0.333,0.334) | 有 | 54.2762 | 4.90976 | 0.24416 | 109790 | 44 |
| 3 | (0.2,0.6,0.2) | 有 | 54.4854 | 4.94935 | 0.35034 | 154130 | 75 |
| 4 | (0.4,0.4,0.2) | 有 | 53.9304 | 4.90493 | 0.30855 | 119137 | 53 |
| 5 | (0.5,0.3,0.2) | 有 | 53.834 | 4.92442 | 0.31335 | 103229 | 42 |
| 6 | (0.2,0.4,0.4) | 有 | 54.3292 | 4.9164 | 0.22891 | 121138 | 52 |
| 7 | (0.3,0.3,0.4) | 有 | 54.0309 | 4.97826 | 0.32397 | 108033 | 45 |
| 8 | (0.4,0.2,0.4) | 有 | 53.9457 | 4.94509 | 0.25647 | 82807 | 29 |
| 9 | (0.6,0.2,0.2) | 有 | 54.4326 | 4.9525 | 0.30684 | 80039 | 29 |
| 10 | (0.4,0.2,0.4) | 有 | 53.9457 | 4.94509 | 0.25647 | 82807 | 29 |
| 11 | (0.2,0.2,0.6) | 有 | 53.7674 | 4.86648 | 0.19939 | 77126 | 28 |
| 12 | (0.4,0.4,0.2) | 有 | 53.9304 | 4.90493 | 0.30855 | 119137 | 53 |
| 13 | (0.3,0.4,0.3) | 有 | 54.1027 | 4.88994 | 0.28864 | 121659 | 54 |
| 14 | (0.2,0.4,0.4) | 有 | 54.3292 | 4.9164 | 0.22891 | 121138 | 52 |
| 15 | (0.2,0.6,0.2) | 有 | 54.4854 | 4.94935 | 0.35034, | 154130 | 75 |
| 16 | (0.2,0.4,0.4) | 有 | 54.3292 | 4.9164 | 0.22891 | 121138 | 52 |
| 17 | (0.2,0.2,0.6) | 有 | 53.7674 | 4.86648 | 0.19939 | 77126 | 28 |

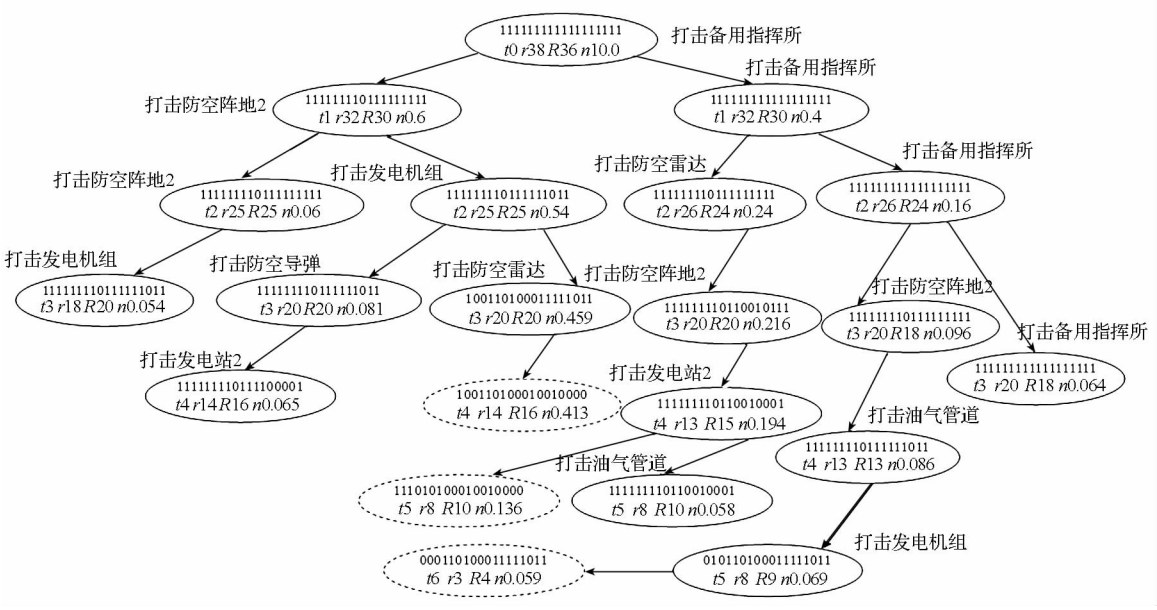


图 4 目标打击策略树

Fig. 4 Target striking tree

从算例 1、2 可以看出,在使用启发式后,求解的策略的统计数据要优于无启发式情况,由于启发式将预期不能到达目标体系失效的状态排除,因此算法访问的状态空间大大缩小,算法运行时间缩短,提高了运行效率,优化了结果策略。

其次,代价权重的调整,使得到的结果策略发生变化,对应的资源消耗、打击阶段数、体系失效概率等指标也发生变化。根据式(7),代价权重的调整使得每个状态下的行动代价值发生变化,根据式(10)可知这会影响算法在解空间的搜索轨迹,进而影响结果策略。

从算例 3 ~ 5 和 6 ~ 8 可知,在失败代价权重不变的情况下,资源代价权重的提高会使打击目标体系策略的总资源消耗减少。

然而,在失败代价权重不变时,时间代价权重的提高和结果策略的时间消耗减少之间没有必然关联。一方面时间权重提高,会使求解过程趋于搜索总阶段数较少的方案;另一方面资源代价权重降低(因为失败代价权重固定),使得求解过程可能搜索消耗资源较多的方案,而单阶段内可消

耗资源有限,因此消耗资源较多的方案对应的总阶段数也较多。两种相反搜索趋势使结果策略时间消耗变化不确定。

从算例 9 ~ 11、12 ~ 14 和 15 ~ 17 得知,当固定时间代价权重或资源代价权重时,失败代价权重的提高,会使得成功概率高的行动被选中可能性提高,因此得到的结果策略的失败概率减少。

最终,权重系数调整,对于结果策略的指标影响幅度较小,说明了算法的稳定性。

计算得到的目标打击策略能够使用树的形式直观描述,其中算例 2 对应的目标打击策略树如图 4。图中的椭圆表示目标体系在被打击过程的某个中间状态,椭圆内的 1、0 排列表示目标体系各要素状态的取值,对应顺序为 P、K、G1、G2、T、E、F、C1、C2、C、B、R、S、D、M1、M2、M、SoS。椭圆中的 $t$ 代表时间状态,$r$ 代表资源 A 剩余数量,$R$ 代表资源 B 剩余数量,$n$ 代表从初始状态到达该状态的概率值,椭圆外侧的注释表示在该状态下所打击的目标单元。图中的虚线椭圆为终止状态(结果状态的出现概率小于 0.01 的不在图中显示)。

## 5　小结

目标选择是军事决策的重要组成部分,针对打击目标体系过程中的目标选择问题,本文使用与或树对目标体系各层状态间影响关系进行了建模,并建立了基于离散时间 MDP 的目标选择模型,提出了基于使目标体系失效的最小资源消耗和失败概率的启发式算法,压缩了 MDP 问题求解的状态空间,提高了求解效率,并用案例验证了该算法的有效性。案例表明,该方法能有效解决具有复杂目标关联的多阶段目标选择策略问题。

## 参考文献(References)


[1] Zhou Y, Zhu C, Lei T, et al. A COG analysis model of system-of-systems (SoS) based on multi-entity Bayesian networks (MEBNs) [C]//The 13th International Conference on Artificial Intelligence (ICAI), Las Vegas, USA, 2011.

[2] 刘健,王献锋,聂成. 空袭目标威胁程度评估与排序[J]. 系统工程理论与实践,2001,21(2):142 - 144.
LIU Jian, WANG Xianfeng, NIE Cheng. Evaluation and sorting of the air targets' threat[J]. Systems Engineering-Theory & Practice, 2001,21(2):142 - 144. (in Chinese)

[3] 吴智辉,张多林. 基于模糊理论的空袭目标威胁判断模型[J]. 火力与指挥控制,2005,30(4):92 - 94.
WU Zhihui, ZHANG Duolin. A model for the air targets' threat evaluation based on fuzzy theory[J]. Fire Control and Command Control,2005,30(4):92 - 94. (in Chinese)

[4] Falzon L. Using Bayesian network analysis to support centre of gravity analysis in military planning[J]. European Journal of Operational Research, 2006, 170(2):629 - 643.

[5] 朱延广,朱一凡. 基于影响网络的联合火力打击目标选择方法研究[J]. 军事运筹与系统工程,2010,24(3):64 - 69.
ZHU Yanguang, ZHU Yifan. Research on targets selecting methods in joint fire strike based on influence net[J]. Military Operations Research and Systems Engineering,2010,24(3):64 - 69. (in Chinese)

[6] Pousi J. Decision analytical approach to effects - Based operations[D]. Helsinki: Helsinki University of Technology, 2009, 35 - 63.

[7] 秦前付,曹存根,徐洸. 基于图论的作战计划军事效果评估[J]. 计算机科学, 2005, 32(7), 148 - 151.
QIN Qianfu, CAO Cungen, XU Guan. Evaluating military effect of air operational plan based on network theory[J]. Computer Science, 2005, 32(7), 148 - 151. (in Chinese)

[8] 李新其,向爱红,李红霞. 系统目标毁伤效果评估问题研究[J]. 兵工学报,2008,29(1):57 - 61
LI Xinqi, XIANG Aihong, LI Hongxia. Calculation and assessment on damage effect of system target[J]. Acta Armamentarii, 2008,29(1):57 - 61. (in Chinese)

[9] 袁震宇,谢春思,张宇,等. 基于故障树的系统目标打击决策模型研究[J]. 舰船电子工程, 2010,30(7):52 - 55.
YUAN Zhenyu, XIE Chunsi, ZHANG Yu, et al. Model study of system target attacking decision based on fault tree[J]. Ship Electronic Engineering, 2010,30(7):52 - 55. (in Chinese)

[10] 蔡怀平,刘靖旭,陈英武. 动态武器目标分配问题的马尔可夫性[J]. 国防科技大学学报,2006,28(3):124 - 127.
CAI Huaiping, LIU Jingxu, CHEN Yingwu. On the Markov characteristic of dynamic weapon target assignment problem[J]. Journal of National University of Defense Technology,2006,28(3):124 - 127. (in Chinese)

[11] 解武杰,冯锦丽. 基于马尔可夫过程的防空武器目标选择[J]. 空军工程大学学报(自然科学版),2009,10(3):37 - 42.
XIE Wujie, FENG Jinli. Anti-aircraft weapons target choice based on Markov process[J]. Journal of Air Force Engineering University(Natural Science), 2009, 10(3): 37 - 42. (in Chinese)

[12] Boutilier C, Dean T, Hanks S. Decision - theoretic planning: structural assumptions and computational leverage[J]. Journal of Artificial Intelligence Research, 1999, 11:1 - 94.

[13] Aberdeen D, Thiebaux S, Zhang L. Decision-theoretic military operations planning [C]//Proceedings of Fourteenth International Conference on Automated Planning and Scheduling, Whistler, 2004:402 - 412.

[14] Meuleau N, Hauskrecht M, Kim K E, et al. Solving very large weakly coupled Markov decision processes [C]// Proceedings of the Fifteenth National Conference on Artificial Intelligence, 1998: 165 - 172.

[15] 阳东升, 张维明, 刘忠,等. 组织过程策略优化的案例分析与求解[J]. 系统仿真学报,2005,17(7): 1648 - 1654.
YANG Dongsheng, ZHANG Weiming, LIU Zhong, et al. Optimizing strategies of organizational processes: Analysis and solutions of case[J]. Journal of System Simulation, 2005,17(7): 1648 - 1654. (in Chinese)

[16] 王长春,陈俊良,陈超,等. 基于复杂网络作战体系破击的建模与仿真[J]. 系统仿真学报, 2012, 24(7): 1491 - 1495.
WANG Changchun, CHEN Junliang, CHEN Chao, et al. Modeling and simulation of combat systems paralysis based on complex network[J]. Journal of System Simulation, 2012, 24(7):1491 - 1495. (in Chinese)

[17] 胡奇英,刘建庸. 马尔科夫决策过程引论[M]. 西安:西安电子科技大学出版社, 2000.
HU Qiying, LIU Jianyong. An introduction to Markov decision process[M]. Xian: Xidian University Press, 2000. (in Chinese)

[18] Barto A G, Bradtke S J, Singh S P. Learning to act using real-time dynamic programming[J]. Artificial Intelligence, 1995, 72(1 - 2):81 - 138.

[19] Bonet B, Geffner H. Labeled RTDP: Improving the convergence of real-time dynamic programming [C]// Proceedings of Thirteenth International Conference on Automated Planning and Scheduling, 2003:12 - 21.