

doi: 10.7690/bgzd.2021.11.020

基于进化行为树的智能体决策行为建模研究

张敬博¹, 吕桐臻²

(1. 杨凌职业技术学院水利工程分院, 陕西 杨凌 712100; 2. 海军葫芦岛试验训练基地, 辽宁 葫芦岛 125000)

摘要: 针对传统仿真系统中建模方法存在领域知识获取困难、生成行为固定、缺乏适应性问题, 提出基于静态约束的进化行为树方法构建智能体决策行为模型。在通用进化算法基础上自动生成反映智能体决策行为逻辑的行为树拓扑结构, 通过通用学习方法与领域规则的较好结合提升决策模型的生成效率和适应性, 并以坦克对战军事游戏中的决策行为建模为例进行验证。结果表明: 该方法能增强决策逻辑的可解释性, 具备可行性和科学性。

关键词: 行为树; 遗传编程; 静态约束; 决策行为建模

中图分类号: TP301 **文献标志码:** A

Research on Decision-making Behavior Modeling Based on Evolutionary Behavior Tree

Zhang Jingbo¹, LYU Tongzhen²(1. *Water Conservancy Engineering Branch, Yangling Vocational & Technical College, Yangling 712100, China;**2. Navy Huludao Experimental Training Base, Huludao 125000, China)*

Abstract: Aiming at the problem of acquiring domain knowledge, fixed behavior pattern and adaptability lack of behavior modeling in traditional simulation systems, this paper establish decision-making behavior model, proposes an evolutionary behavior tree method based on static constraints. Based on the general evolutionary algorithm, the behavior tree topology structure that reflects the decision-making behavior logic is automatically generated, and the generation efficiency and adaptability of the decision-making model is improved through combination of general learning methods and domain rules. Preliminary experiments in tank game validate the proposed method. The results show that the method enhances the interpretability of decision-making logic, and is feasible and scientific.

Keywords: behavior tree; genetic programming; static constraint; decision-making behavior modeling

0 引言

当前多数作战仿真系统的兵力实体采用有限状态机、规则脚本等知识工程方法建模决策行为^[1], 其本质任务是基于平台或人员的物理行为能力, 对作战场景中操控平台人员的认知决策行为建模。如: 美军 EADSIM 系统采用规则集描述不同作战单元的战术规则^[2]; 文献[3]中, 坦克战车计算机生成兵力(computer generated forces, CGF)通过有限状态机实现对行为决策的处理; 文献[4]基于 Petri 网构建舰船的反潜战术模型。以上方法能够按专家知识灵活定制行为、可理解性强, 但存在领域知识获取困难、建模工作量大、效率低等问题, 且产生的行为固定、缺乏适应性。

基于机器学习的建模方法能够利用学习技术从经验数据中自主生成决策行为模型^[5]。其中进化算法是一类在 CGF 行为建模中被广泛采用的学习技术。A. TAYLOR 等采用进化算法对军事训练和实验

中的 CGF 行为生成进行研究, 并在简单的游戏虚拟场景中与手工编码的控制器进行对比^[6]。L. J. LUOTSINEN 等采用遗传编程生成追逃游戏中的 CGF 行为, 并与人为开发固定脚本进行对比^[7]。进化算法不关注推理决策过程, 强调行动的输出, 主要通过与环境交互学习进行知识纠错, 提高应对复杂未知情况的适应能力^[8]。由于机器学习的黑盒性质使得模型可解释性差, 加之学习在复杂条件下的收敛性等问题可能导致行为不可控, 使得模型可信度降低。

针对上述不足, 笔者提出考虑静态约束的进化行为树算法构建军用仿真或游戏中单兵、平台等智能体的决策行为模型。在通用进化算法基础上, 以行为树为模型表示方法, 以人员或平台的物理行为模型和感知态势条件为叶子节点, 自动生成反映智能体决策行为逻辑的行为树拓扑结构, 增强模型的可解释性。同时, 通过设计行为树语法模式和军事领域规则作为静态约束, 减小解搜索空间,

收稿日期: 2021-06-27; 修回日期: 2021-08-20

基金项目: 湖南省自然科学基金(2016JJ4006)

作者简介: 张敬博(1984—), 男, 陕西人, 硕士, 从事数值模拟研究。E-mail: 309462032@qq.com。

提升算法的学习效率和生成模型的质量。最后，在 Robocode 坦克对抗仿真实验中，证明了该方法进化得到的最优个体在具备可解释性的同时，相比人工构建的行为树模型具有更高的适应性。

1 静态约束进化行为树建模方法

行为树是具有根节点的有向树结构，一棵行为树通常包括序列、选择等控制节点和条件、动作等叶子节点，通过层次化方式将复杂任务进行分解，用模块化方式描述智能体任务逻辑之间的切换^[9-10]。

进化行为树算法用于智能体决策行为建模，一般采用遗传编程优化随机生成的行为树种群，以获得满足评估指标要求的模型。每个行为树个体表示一个候选控制智能体的决策行为模型，其合理性由专家评估准则确定的适应度函数来衡量。在此基础上，智能体通过与仿真环境的试错交互优化专家评估函数，寻找最高适应值的行为树个体，将其作为最终生成的决策模型。然而由于行为树语法规则非常灵活，仅依靠适应度函数和无限制的进化算子引导，庞大的随机搜索空间和任意交叉带来的解膨胀问题^[11]，往往导致学习效率低、生成解效果不好。

笔者提出的基于静态约束进化行为树的决策行为建模方法，主要通过种群初始化和进化过程中对行为树结构设计引入静态约束，以提高行为树生成的效率和解质量。

1.1 静态约束设计

针对行为树决策建模，笔者设计通用和领域应用 2 类静态约束，一是行为树语法模式对结构的约束，二是领域规则对行为树结构的约束。

1.1.1 行为树语法规则约束

行为树的语法规则模式可在不改变其执行逻辑

的情况下简化冗余的组合，笔者从行为树基本语法出发，设置以下规则减小搜索空间以加快学习。

- 1) 设置根节点为选择节点；
- 2) 上下 2 层的控制节点设置为不同类型，即上层控制节点为选择节点，则下层控制节点应为序列节点或并发节点；
- 3) 限制控制节点下的条件数量不超过 2 个，其后可跟随 1 个或多个动作节点；若只有 1 个终止节点，则必须为动作节点。

规则 1)和 2)约束典型的与或树结构，避免无效组合；规则 3)按照常用的行为树态势处理形态反映行为树决策过程，可以视为不断地进行子态势处理。

1.1.2 领域规则约束

为实现智能体行为的逼真性和智能性，使决策模型产生的行为逻辑能够满足相应领域规则的约束，需要在决策行为建模过程中将领域规则考虑在内。领域规则对行为树结构的约束主要表现在：

1) 更改或删除行为树语法无误，但不符合实际应用基本规则的子树，以保证最终得到的行为树模型满足规则约束。如终止节点的组合，不能出现行为上的矛盾；前进和后退 2 种行为不能同时作为并行节点的子节点；坦克等平台在同一态势下决策判断不超过 2 次，树的最大深度不超过 6 层等。

2) 平台或人员在战场上的行为，通常可认为是任务目标不同态势下对战术机动、侦察、攻击等行动的处理；因此，在行为树进化前，可根据领域规则形成基本的决策行为框架，将平台的主要决策行为由 1 个序列节点控制，下面由几个相互独立的战术行动处理事件组成。

1.2 算法框架流程

基于上述静态约束，采用进化行为树方法对智能体决策行为建模，框架如图 1 所示。

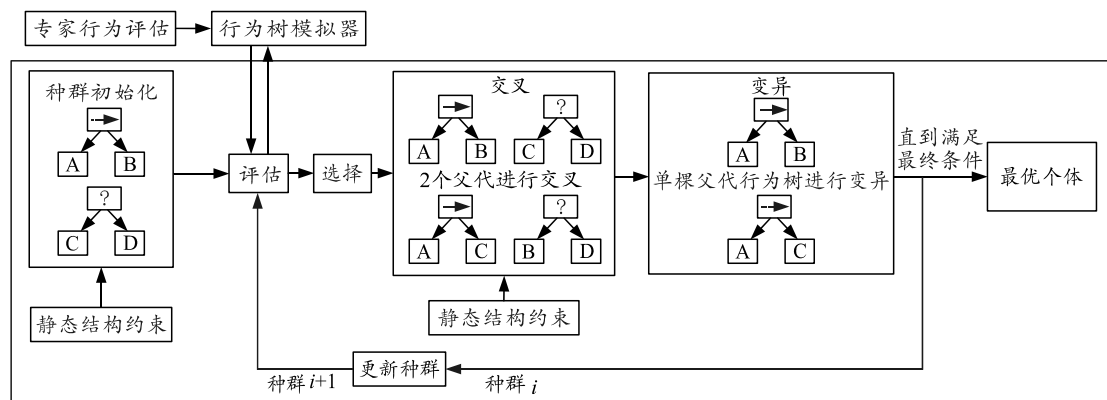


图 1 基于静态约束的进化行为树的决策模型框架

框架主要包括决策任务的领域知识建模、行为树决策模型仿真器、基本的进化算法模块以及静态结构约束组件。决策任务的领域知识建模要求建模人员根据具体任务,确定决策相关的基本状态、原子动作及其他领域知识。行为树决策模型仿真器,对所生成的行为树决策模型进行模拟评估。仿真器模拟智能体所执行作战任务,智能体由种群个体所表示的行为树控制,决策模型的优劣由领域专家提供的任务性能指标衡量。“静态结构约束”组件能避免生成许多无意义和低效的树结构。算法过程为:

1) 种群初始化:遗传编程系统依据静态约束生成一定数目的初代行为树种群。

2) 种群评估:针对种群中的每一个体行为树,将其作为仿真模型中任务实体的控制策略,多次运行仿真,根据专家评估函数计算个体的适应度。

3) 个体选择:遗传编程系统从当前代的行为树种群按照一定策略,选择较高适应度的个体行为树用于进一步的遗传操作。

4) 遗传操作:对被选择的个体执行交叉和变异操作,执行交叉操作时需遵守静态约束,生成新的行为树个体加入到子代。

5) 遗传操作后,更新种群到下一代继续进化,直至满足进化结束的最终条件,适应值最高的行为树即为所生成的行为树模型。终止条件通常包括达到最大进化代数或一定代数内最高适应度不再变化。

2 仿真实验

为了评估所提出算法在军用仿真和游戏中智能体决策行为模型中的表现,笔者选择 Robocode 游戏对抗平台作为算法验证的仿真测试平台,Robocode 是一款坦克机器人游戏对抗平台。在仿真环境中,坦克智能体具有能量值、速度、炮弹类型、炮管冷却速度 4 个主要属性。坦克被命中、碰撞或发射炮弹都会损失一定能量,能量为 0 判定为毁伤;炮弹类型可变,用来模拟真实坦克火力系统不同武器的杀伤力。任务中,双方坦克智能体需要根据感知信息在给定地图环境中进行自由交战,目标是尽快尽量多地击败敌方坦克并保存自己,取得胜利。

仿真实验中,以手工生成的行为树决策模型为基准,对学习过程性能和最终生成的决策模型进行了对比分析。实验的主要目标是验证所提方法相比人工行为树建模,是否能够在保持决策逻辑可解释基础上具备更优的学习效果。

2.1 实验建模

对坦克智能体的决策行为进行建模,本质是对操控坦克人员的决策控制逻辑建模,基础是战场态势感知和坦克的移动、开火等原子物理行为模型。采用进化行为树生成决策模型时,其函数节点包括序列节点和选择节点,终止节点包括坦克智能体相关的不同的条件节点和动作节点。根据对仿真平台的分析,设置用于进化的行为树基础节点如下:

1) 条件节点是智能体对战场态势认知的具体体现,战场态势中最为重要的就是敌方兵力;所以,条件节点主要集中于对抗游戏中智能体对敌信息的获取。条件节点包括:

自身被命中:坦克自身是否被敌方炮弹命中。若被命中,则返回 true。

锁定目标:火控系统在战场侦察中锁定敌方坦克等目标,则返回 true。

敌人发起攻击:在锁定敌方坦克等目标的情形下,若感知到敌方发起攻击,则返回 true。

敌人未攻击:在锁定敌方坦克等目标的情形下,若感知到敌方未攻击,则返回 true。

击中敌方坦克:在对敌方进行进攻时,若成功击中敌方坦克,则返回 true。

击中障碍物:在对敌方进行进攻时,若击中分布在战场上的障碍物,则返回 true。

弹药充足:坦克炮弹数量大于 0,则返回 true。

2) 动作节点反映坦克基本的物理行为能力,主要结合对抗游戏中的坦克物理模型,描述坦克在机动、火力、侦察方面的原子行为。动作节点包括:

战场侦察:战场侦察是坦克对周围区域进行感知,以发现、识别敌方坦克,可以配置旋转角度参数,取值范围为 $0\sim 360^\circ$ 。

发射炮弹:坦克发射炮弹,可以配置炮弹种类参数,取值为 1、2、3 不同能量的弹种。

机动:坦克前后的基本物理移动,可以配置机动距离参数,取值范围为距离大于 0。

车体转动:坦克车体向左或向右转动,可以配置转动角度参数,取值范围为 $0\sim 360^\circ$ 。

炮塔转向:坦克炮塔向左或向右转动,可以配置转动角度参数,取值范围为 $0\sim 360^\circ$ 。

火炮转动:火炮向左或向右转动,可以配置转动角度参数,取值范围为 $0\sim 360^\circ$ 。

3) 控制节点,即为行为树的函数节点,包括序列节点和选择节点 2 种基本类型。

4) 评估准则：决策优化的评估目标设定为在不同战场态势下通过综合设置机动、火力运用、战场侦察策略，使得在不同战场都能以相对小的代价毁伤相对多的敌人。设击毁一辆敌方坦克得 1 分，自身损毁得-1 分，每回合统计评分之和为本方对抗的总分值，将模型适应度设为本方评分与对抗双方总分的比值。模型个体最终的适应度为多回合对抗后的适应度均值。

5) 静态约束：进化过程中，所设计的通用行为树语法规则约束遵循 2.1.1 节，根据领域规则约束将坦克智能体行为按照侦察发现敌方、意外遇敌攻击、触碰障碍等典型态势决策框架设计，基本结构如图 2 所示。

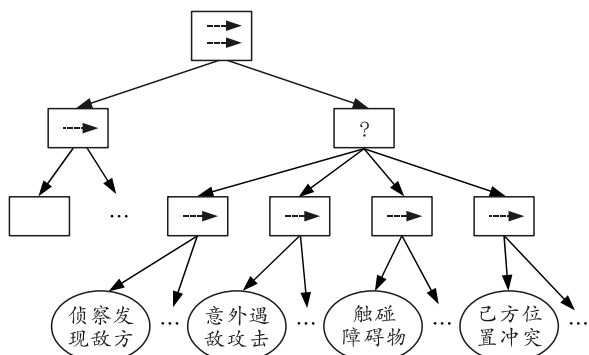


图 2 静态约束下的坦克智能体行为树结构

2.2 参数设置

如表 1 所示，实验中进化行为树生成的模型对战对手包括平台中自带的代号为 Crazy、Fire、RamFire 的 3 辆坦克。

表 1 仿真平台自带坦克模型介绍

坦克	战斗模式简介
Fire	360°不断侦察搜索，发现敌方即多种模式攻击
Crazy	大角度曲线运动并同步搜索，发现敌方即多种模式攻击
RAMFire	随机运动接敌后立即开火

手动创建行为树遵循的战术规则：未发现敌方坦克时，开启搜索模式寻找敌方；若发现敌方，则根据敌方所处位置转向后进行攻击；若被敌方攻击，则执行撤退。进化行为树算法的种群大小、进化代数、交叉概率等参数设置如表 2 所示。

表 2 进化行为树算法参数设置

种群大小	进化代数	交叉概率	变异概率	精英数量	对战回合
40	51	0.05	0.1	10	10

所提算法根据上述参数，以图 2 所示结构为基础，生成一定数目的初代行为树种群，选择每一代的优势个体经过复制进入下一代，非优势个体依据静态约束经遗传变异操作后进入下一代评估，最终

选择最优的完备行为树逻辑作为生成的决策模型。

2.3 结果分析

在学习过程中，笔者记录了每一代所有个体的适应度，学习结束后生成最优个体的测试结果。对所得结果进行对比分析，重点分析最终学习效果以及所生成模型产品的形态。

1) 模型生成时间及可读性分析。

首先比较手工构建与进化行为树算法生成行为树模型所用时间。手工构建一个完整的行为树模型，需要领域专家、编程人员反复互动，调研平均需要 1 周时间。利用进化行为树算法生成最优行为树，在笔者的实验中从领域知识设计、算法的节点设计到学习完成最多仅需 3 天。以上仅基于较为简单的坦克攻防场景，若将进化行为树算法应用于更加复杂的场景，由于手动构建行为树越发困难，进化行为树算法的模型生成效率将成为巨大优势。

图 3 展示了经过静态约束进化行为树算法学习后得到的最优个体行为树模型，可以看到模型将完整状态空间的决策划分为一些特定的情况来处理。例如，当发现敌人时，智能体选择合适的炮弹开火、转向及机动等动作；触碰到障碍物时，选择转向和连续的机动以避免。这种以人类可理解的方式所组织的决策模型是可读可解释的；同时，又能够完成人类所不擅长的具体参数细节确定，发现一些人类未察觉的动作序列组合，可以被视为计算创造力。

2) 学习生成的模型效果对比。

将进化过程中最优个体与手工行为树及平台自带的 3 种模型进行对抗实验，查看最终学习效果与手工开发模型的决策目标满足度。如表 3 所示，首先，将进化行为树算法第 0、20、35、51 代优势个体分别与平台自带的 3 种坦克模型进行对抗，记录评估值。其次，将手动构建行为树模型与平台坦克模型、进化个体分别对抗，并记录评估值。从进化过程不同代数最优个体之间的对比，可以看出随着行为树的不断迭代进化，所得到的模型性能不断提升，证明了种群学习进化的有效性。从手工构建的行为树模型与进化过程最优个体对抗的评估值可以看出：专家构建的行为树模型能力相当于 20 代后的进化个体，进化行为树算法能够生成更好的决策行为模型。

从进化过程不同代数最优个体之间的对比，可以看出随着行为树的不断迭代进化，所得到的行为树模型性能不断提升，证明了种群学习进化的有效

性。从手工构建的行为树模型与进化过程最优个体对抗的评估值可以看出：专家构建的行为树模型能

力相当于 20 代后的进化个体，进化行为树算法能够生成更好的决策行为模型。

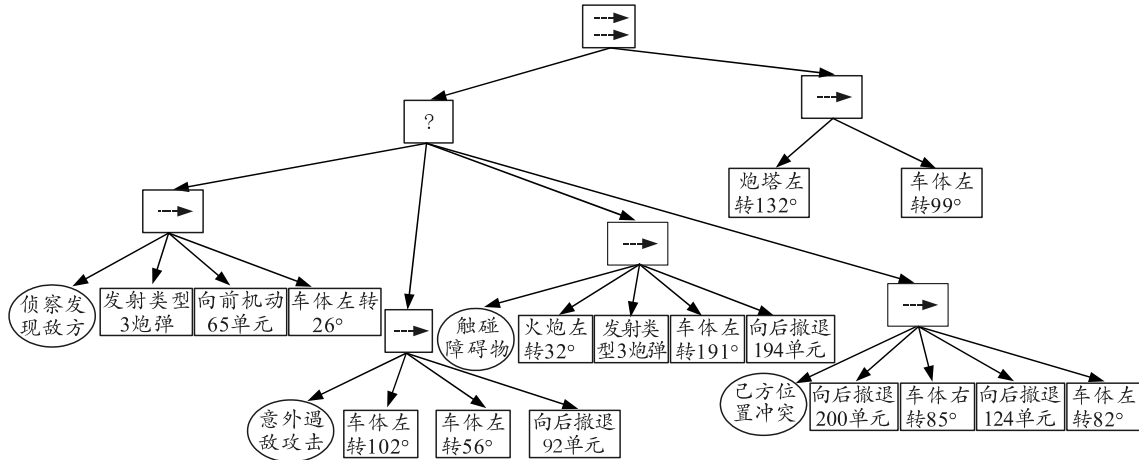


图 3 最佳个体的行为模型

表 3 “我”方对抗评估值

行为树模型	Fire	Crazy	RamFire	评估值	0	20	35	51
第 0 代最优	0.02	0.11	0.11	0.12		0.14	0.04	0.03
第 20 代最优	0.09	0.42	0.03	0.27	0.86		0.05	0.17
第 35 代最优	0.60	0.67	0.33	0.53	0.96	0.95		0.37
第 51 代最优	0.75	0.79	0.54	0.69	0.97	0.83	0.63	
手工行为树	0.58	0.15	0.20	0.31	0.90	0.43	0.13	0.09

3 结论

笔者提出基于静态约束进化行为树算法来构建作战仿真和游戏当中虚拟坦克智能体的决策行为模型，并基于 Robocode 坦克机器人对抗平台对模型运行效果进行仿真实验。该方法将可解释的行为树表示方法与全局优化的进化学习方法相结合，并且基于智能体的静态约束缩小通用行为树进化的学习空间，提升学习效率，改进学习效果。实验验证了该方法进化得到的优势个体具有更高的适应性，证明了种群学习进化的有效性。在此基础上，考虑如何将模仿学习与行为树结合以获得反映特定仿真对象行为逻辑的决策模型，是下一步的研究方向。

参考文献：

[1] 杨瑞平, 黄晓冬. 计算即生成兵力智能决策方法及其应用[M]. 北京: 电子工业出版社, 2011: 8.
 [2] BOURASSA N R. Modeling and Simulation of Fleet Air Defense Systems Using EADSIM[D]. Thesis Collection. USA: Naval Postgraduate School, 1993: 5.
 [3] 李一. 基于 Unity3D 的虚拟战场环境研究与实现[D].

石家庄: 河北科技大学, 2017: 2.

[4] MURATA T. Petri nets: Properties, analysis and applications[J]. Proc IEEE, 2010, 77(4): 541-580.
 [5] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 5.
 [6] TAYLOR A. Designing Controllers for Computer Generated Forces with Evolutionary Computing: Experiments in a Simple Synthetic Environment[R]. Canada: DRDC, 2013.
 [7] LUOTSINEN L J, KAMRANI F, HAMMAR P, et al. Evolved creative intelligence for computer generated forces[C]. In Systems, Man and Cybernetics (SMC), 2016 IEEE International Conference on, 2016: 3063-3070.
 [8] ZHANG Q, YAO J, YIN Q, et al. Learning Behavior Trees for Autonomous Agents with Hybrid Constraints Evolution[J]. Applied Sciences, 2018, 8(7): 1077.
 [9] 刘晓伟, 高春鸣. 结合行为树与 Q-Learning 优化 UT2004 中 agent 行为决策[J]. 计算及工程与应用, 2016, 52(3): 113-118.
 [10] SEKHAVAT Y A. Behavior trees for computer games[J]. International Journal on Artificial Intelligence Tools, 2017, 26(2): 1-27.
 [11] POLI R, LANGDON W B, MCPHEE N F. A Field Guide to Genetic Programming[J]. Genetic Programming & Evolvable Machines, 2009, 10(2): 229-230.