

DOI: 10.19650/j.cnki.cjsi.J2210593

基于驾驶员模型的六足机器人自主/协同决策*

陈潇磊^{1,2}, 尤波^{1,2}, 李佳钰^{1,2}, 丁亮³, 董正¹

(1. 哈尔滨理工大学黑龙江省复杂智能系统与集成重点实验室 哈尔滨 150080; 2. 哈尔滨理工大学先进制造智能化技术教育部重点实验室 哈尔滨 150080; 3. 哈尔滨工业大学机器人技术与系统国家重点实验室 哈尔滨 150001)

摘要:重载六足机器人在野外地形环境移动作业时的决策智能水平亟待提高。然而,当机器人在尚未形成合理的决策结构层次时,直接采用其与环境进行交互方式进行常规的强化学习训练,将导致机器人的行为决策过于发散。因此,本文首先利用一种符合驾驶员决策逻辑的分步训练神经网络,得到驾驶员的决策经验模型,使机器人快速形成自主决策智能。此外,为融合人机决策优势,本文基于合作博弈理论,提出一种消除人机协同决策指令冲突的方法。搭建面向重载六足机器人人机协同决策的半物理仿真实验系统,开展实验的结果表明,机器人通过学习驾驶员先验模型和自主训练,其决策效果可接近驾驶员决策水平,同时人机协同决策指令可有效弥补单智能体决策指令的缺陷,在规则沟壑地形下协同决策指令的碰撞率指标优于驾驶员单智能体指令 23.8%,障碍地形下协同决策指令的能量消耗指标优于机器自主单智能体指令 34.1%。

关键词: 六足机器人; 协同决策; 驾驶员先验模型; 半物理仿真; 神经网络

中图分类号: TP24 TH39 **文献标识码:** A **国家标准学科分类代码:** 510.80

Hexapod robot self/collaboration decision based on the driver's prior model

Chen Xiaolei^{1,2}, You Bo^{1,2}, Li Jiayu^{1,2}, Ding Liang³, Dong Zheng¹

(1. Heilongjiang Provincial Key Laboratory of Complex Intelligent System and Integration, Harbin University of Science and Technology, Harbin 150080, China; 2. Key Laboratory of Advanced Manufacturing Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China; 3. State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China)

Abstract: The level of decision-making intelligence of heavy-duty hexapod robots in the field terrain needs to be improved. However, if robots have not yet formed a reasonable decision structure level, the conventional decision-making reinforcement learning which is directly interact with the environment, will lead to the robot's decision-making being too divergent. Therefore, this article first obtains the driver's decision-making experience model through a step-training neural network which conforms to the driver's decision-making habits. Hence, the robot can quickly form decision-making intelligence. In addition, to better play the advantages of human-robot decision-making, this article proposes a method to eliminate the conflict of human-robot coordinated decision-making commands based on the cooperative game theory. A semi-physical simulation experiment system for human-machine collaborative decision-making of heavy-duty hexapod robots is designed and established. After carrying out experimental verification around the proposed methods, results show that the robot can approach the driver decision-making effect by learning the driver's prior model and reinforcement training, and the effect of the human-robot collaborative decision-making commands can also make up for the defects in unilateral decision-making. In the regular ditches terrain, the collision index of the collaborative decision commands is 23.8% better than that of the single driver agent commands; in the obstacle terrain, the energy consumption index of the collaborative decision commands is better than that of the single robot agent commands by 34.1%.

Keywords: hexapod robot; collaborative decision; driver's prior model; half-physical simulation; neuro network

0 引言

重载六足机器人因其地形通过性强,承载能力大,被广泛应用于山地丛林^[1]、星球表面^[2]、水下^[3]等复杂未知地形环境下执行移动作业任务。为减轻其自身结构复杂性带来的驾驶操纵负担,其运动行为决策方式由早期的驾驶员全程手动驾驶模式,逐渐转为人机协同驾驶决策模式^[4-5]。然而,二者可以在决策层进行更深度协同的前提是机器人自身的决策智能水平需接近驾驶员决策的水平。因此,如何提高足式机器人在野外地形环境移动作业时的自主决策智能水平及人机协同能力,具有较大研究意义和应用价值。

在利用深度神经网络或强化学习的构架,训练足式机器人在非平坦地形下的通过能力方面,已经有很多机构做出了卓越的成果。其中最具影响力的是苏黎联邦理工大学研发的四足机器人 ANYmal^[6],研究者利用强化学习方法对机器人各关节电机的控制参数进行训练,并将训练结果迁移至实物样机,使之适应更加复杂多变的室外地形环境,其以小步快跑步态下楼梯的成功率达到了70%以上^[7]。上述方法无疑对小型轻量化的四足机器人是有效的,其成功避免了人类操作者对机器人示教训练的冗长过程,机器人可以在强化学习的框架下通过与环境交互,进行大量试错-纠正的循环迭代,最终以机器的数据计算优势获得预期的训练结果。然而对于体积较大的重载六足机器人,尤其是对于需要拓展出载人功能的六足机器人,仅依靠机器人在环境中以试错-纠正为内核的训练,无论是对机器人的本体结构还是驾驶员的安全都存在不可忽视的缺陷。

不同于人类操作者的单步示教方法以及直接将机器人放在环境中进行试错学习的方法,目前,越来越多的学者聚焦于归纳量化包括人类在内的足式动物的经验模型,并将其应用于提高足式机器人运动智能水平。在此领域最具代表性的研究成果是由 Cully 等^[8]将腿部受伤狗的运动经验归纳成行为空间集合,指导运动能力缺失的六足机器人进行容错运动控制。考虑到人类驾驶员在对六足机器人的驾驶决策已经积累了丰富的经验,但驾驶员决策过程是一个复杂的非线性多输入/输出过程,虽然可以尝试以传统的神经网络结构对此类问题进行归纳建模^[9-11]。但其黑箱特性使得输出的决策结果不具有可解释性,导致无法根据驾驶员的决策逻辑习惯对其进行分析和优化,这也是目前该领域亟待解决的问题之一。

此外,随着越来越多先进的人机协同操控系统被开发出来^[12-14],人们开始对与人进行合作的机器人智能水平有着迫切的需求。更值得注意的是,随着机器

自主程度的提高,决策系统不应再简单地执行驾驶员的指令,而是迫切需要找到一种合理化解人指令冲突的方法,将人机的决策指令进行有效地共享与合成^[15-16]。因此,本文研究内容聚焦于归约驾驶员先验决策模型的同时保留驾驶员的决策逻辑流程,进而使机器人带着先验模型去开展自主的决策训练以及与驾驶员协同决策。

1 机器人运动学特征与驾驶员决策逻辑特点

1.1 重载六足机器人结构与运动学特征

本文所述的重载六足机器人 ElSpider I,为仿昆虫结构的电驱动机器人,如图1所示。机器人主要由机体和6个腿部部件构成。机器人的主体结构为六棱柱形状,因其结构对称的特点,使得机器人可以较为容易地在各个方向进行必要的运动。为了有效避免多个腿之间有可能发生的运动干涉问题,6个相同的腿独立且均匀分布在六棱柱形状机体的6个侧面上。每条腿都有3个关节自由度:基关节、髌关节和膝关节。并分别建立世界坐标系、机器人坐标系、基关节坐标系、髌关节坐标系、膝关节坐标系和足端坐标系。

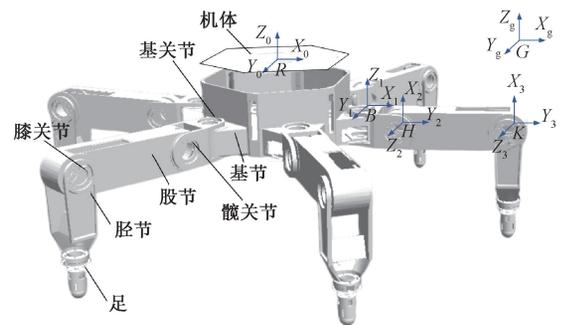


图1 六足机器人构型

Fig. 1 Structure of the hexapod robot

值得注意的是,机器人的自重达到约530 kg,其最大载重可达155 kg,所以该机器人是一款总重约达0.7 t,具有18个自由度的重载六足机器人。若采用驾驶员对其逐一控制的方式,不仅会为驾驶员带来沉重的驾驶负担,并且整机的运动协调性也将大大降低,因此,驾驶员及机器人决策智能应主要关注于六足机器人运动行为的顶层决策,而对于底层运动控制,应根据机器人运动学特征建立理论公式,实现在底端执行层对位置和速度的自主解算与运动控制。

机器人各关节驱动电机均相同,设跟关节转角为 α ,髌关节转角为 β ,膝关节转角为 γ 。基节长度为 L_1 ,大腿(股节)长度为 L_2 ,小腿(胫节)长度为 L_3 ,机体中心距地面高度为 H ,步长为 S_l ,基于上述参数可推导出单腿正运

动学表达式为:

$$\begin{cases} {}^0P_{F_x} = L_1 c \alpha + L_2 c \alpha c \beta + L_3 c \alpha c (\beta + \gamma) \\ {}^0P_{F_y} = L_1 s \alpha + L_2 s \alpha c \beta + L_3 s \alpha c (\beta + \gamma) \\ {}^0P_{F_z} = L_2 s \beta + L_3 s (\beta + \gamma) \end{cases} \quad (1)$$

指定 ${}^0P_F = ({}^0P_{F_x}, {}^0P_{F_y}, {}^0P_{F_z})^T$ 代表足端在 $\Sigma_{0-x_0y_0z_0}$ 坐标系下的位置向量,且 c 代表 \cos , s 代表 \sin 。同理,单腿逆运动学的表达式容易推导,本文不再展开。

六足机器人行走过程中可以被看作是与地面构成的闭合运动链,因此机体与足端的位置关系可表示为:

$${}^C P_F = {}^C P_R + {}^C R_R {}^R P_F \quad (2)$$

其中, ${}^C P_F$ 代表全局坐标系下足端的位置, ${}^C P_R$ 代表全局坐标系下机体的位置, ${}^C R_R$ 代表全局坐标系与机体坐标系间的旋转变换矩阵, ${}^R P_F$ 代表机体坐标系下足端的位置。

上式两边对时间进行求导,结合六足机器人的单腿运动学模型,可得整机的运动学模型:

$$\dot{q}_s = J^{-1} {}^C R_R^T (-V_B - {}^C \dot{R}_R {}^R P_F) \quad (3)$$

其中, q_s 是腿部关节的转角向量, $J(q_s)$ 是腿部雅可比矩阵。

1.2 驾驶员顶层决策逻辑与机器人底层控制流程

对于面向足端和关节的控制任务,一般由机器人的底层运控算法完成,其控制流程由上述各公式可知,式(1)单腿正运动学及其逆运动学表达式,可以在关节转角与足端位置之间建立双向的转化通路,依据式(2)坐标变换表达式可以在足端位置与机体位置间建立双向的转化通路,二者共同为机器人在运动控制底层实现自主位置解算的理论基础;依据式(3)可以在关节转速和机身质心速度之间建立双向的转化通路,为机器人在运动控制底层实现自主速度解算的理论基础。

不同于机器人一般执行的底层控制任务,驾驶员倾向于在顶层对六足机器人运动行为进行操控决策,其决策逻辑为优先规划整机质心的运动,即从整机下一阶段的目标位置和采用的行走速度两个主要方面做出判断决策,进而再针对步态模式、步幅、步高等步态参数下达指令。而对于关节和足端的底层运动控制任务,则多交由机器人进行控制管理,驾驶员仅保留干预权限,根据实时行走状态在必要时进行微调。

因此,根据上述驾驶员/智能体的行为决策逻辑习惯和人机任务分工特点,本文开发一套适用于驾驶员/智能体决策的操控硬件及界面。其中两个手柄用于输出驾驶员对于目标位置和行走速度的决策判断结果,以及步态参数指令;触摸显示屏用于监控和干预机器人关节和足端的控制状态。驾驶员的环境输入来自于两部分,一部分来自于触控屏中人机界面所显示的数

据信息,另一部分来自于显示屏中六足机器人在地形环境中的实时图像。该套操控硬件在保证功能完备性的前提下,能很好的与上述的人机决策逻辑和控制任务相契合。

2 基于驾驶员先验模型的强化学习方法

2.1 面向驾驶员先验模型的分步训练神经网络结构

虽然目前在对复杂非线性决策问题建模方面,以深度神经网络为代表的机器学习工具具有天然的优势,但对于保留决策者原有决策过程的顺序与逻辑性,受限于当前深度学习采用前端到端模式的黑盒特性,一直是该领域的一个难题。然而当开发者难以得知利用机器学习获得决策模型的AI系统做出某个决策的逻辑依据时,这种决策结果的不可解释性导致难以进行有针对性的调整和优化。尤其值得注意的是,面向涉及重型装备和人员生命安全的领域,AI系统决策过程的逻辑性和决策结果可解释性显得更为必要。

不同于目前利用替代模型法、样例驱动法、探针法等,在模型获得后去解析其可解释性的方法。本文倾向于保留驾驶员对六足机器人进行驾驶决策过程的决策顺序性和逻辑性,将其融入到神经网络结构的设计中,使其在源头具有可解释性。

具体的,本文根据驾驶员习惯先对目标位置和行走速度做出规划判断再发布步态指令的决策逻辑特点,将驾驶员决策神经网络模型的隐含层分为两大功能层次,即判断层和指令层。其中判断层针对机体质心下一阶段的运动目标,设计包括3个子功能层次,即方向判断层、距离判断层和速度判断层,判断层各子功能层并行连接。同时,指令层针对机体质心下一阶段的运动指令,指令层也包括3个子功能层次,即步态/步频指令层、步幅/高指令层和机体位姿指令层,指令层各子功能层并行连接,判断层和指令层串行连接。神经网络的输入层为一个局部地形特征矩阵的所有元素,该矩阵描述的是机器人以其两倍机身长度范围内的地形起伏特征。因此,环境信息以局部地形特征矩阵的形式从输入层先流入判断层后,得出方向、距离和速度的决策判断结果将继续流入指令层,指令层将决策指令序列发送至输出层,最后得到该局部地形下的所有决策指令序列,如图2所示。

2.2 基于先验模型更新的自主强化学习方法

为了提高机器人获得驾驶员决策的先验模型后,其决策的泛化能力。本小节在驾驶员先验模型的基础上,借助以部分可察马尔科夫决策模型为内核的Q学习算法建立先验决策模型的更新机制,开展机器人与环境的交互实验,并通过建立机器人接收环境反馈的奖励方程,以

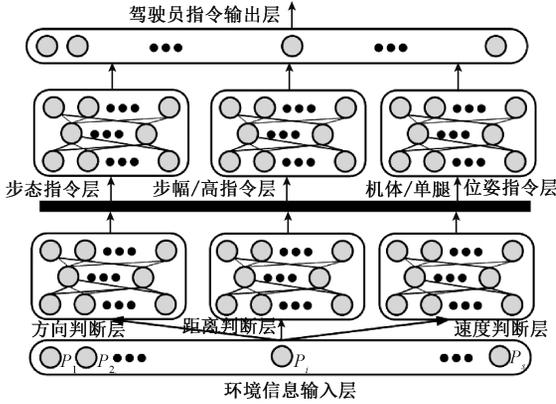


图2 驾驶员决策神经网络模型结构框图
Fig. 2 The structure of the driver decision neural network model

累积奖励最大为原则,对驾驶员的先验决策模型进行更新迭代。

根据前期的实验中执行 a_t 指令序列后,被系统观察并执行的概率过程表明,驾驶员相对于机器智能的决策优势主要体现在稳定性,因此本部分采用改进的马尔科夫模型,将连续的运动行为离散化,并利用稳定裕度指标对机器人与环境交互过程中产生的指令动作进行评价反馈,从而将驾驶员先验模型转化为一种适应环境和自身特征的决策模型。

具体的,机器人以寻找到使奖励方程为最大的指令为目标:

$$\arg \max_{a_t} EU_t(b_t) \quad (4)$$

奖励方程 $EU_t(b_t)$ 由即时奖励和累计奖励两项构成,其表达式为:

$$EU_{t+1}(b_{t+1}) = R(b_{t+1}) + \gamma \max_{a_t} \sum_{o_t} P(o_t | b_t, a_t) \times EU_t(o_t | a_t, b_t) \quad (5)$$

其中, $\overline{S}_m(b_t)$ 代表驾驶员利用离线驾驶系统执行行走任务的平均稳定裕度,将稳定裕度因素纳入即时奖励方程 $R(b_t)$:

$$R(b_t) = \begin{cases} \frac{2}{1 + e^{\frac{S_m(b_t) - \overline{S}_m(b_t)}{S_m(b_t) - \overline{S}_m(b_t)}}} - 1, & S_m(b_t) \leq \overline{S}_m(b_t) \\ \frac{2}{1 + e^{\frac{\overline{S}_m(b_t) - S_m(b_t)}{S_m(b_t) - \overline{S}_m(b_t)}}} - 1.2, & S_m(b_t) > \overline{S}_m(b_t) \end{cases} \quad (6)$$

$$P(o_t | b_t, a_t) = \sum_{s_t} O(o_t | s_t) \sum_{s_{t-1}} T(s_t | s_{t-1}, a_{t-1}) \times b_{t-1}(s_{t-1}) \quad (7)$$

累计奖励方程中,考虑在机器人信任后的 b_t 状态下执行 a_t 指令序列后,被系统观察并执行的概率。传统的

强化学习仅考虑机器人不断与环境间进行交互试错,导致机器人的行为发散,学习模型不容易收敛。本节机器人可参考驾驶员的先验决策行为,其被描述为状态 s_{t+1} , 包含环境特征信息 (x_t, y_t, h_t) 和驾驶员的先验决策指令 a_t^h , 机器人依据强化学习特点在可行指令集中带有一定随机性地生成决策指令 a_t^r 。进一步的,若机器人指令执行结果的碰撞次数低于或高于该地形下的驾驶员历史平均碰撞次数 P_0 , 那么将替换或保留驾驶员的先验决策指令,形成机器人信任的 b_t 状态,其递归表达式可表示为:

$$b_{t+1}(s_{t+1}) = bO(o_{t+1} | s_{t+1}) \sum_{s_t} T(s_{t+1} | s_t, a_t) b_t(s_t) \quad (8)$$

$$a_t = \begin{cases} a_t^h, & P \leq P_0 \\ a_t^r, & P > P_0 \end{cases} \quad (9)$$

2.3 六足机器人决策训练方法

综合 2.1 和 2.2 节可知,为避免机器人在尚未形成合理的决策结构层次时,直接采用其与环境进行交互的方式进行自主决策学习,导致机器人的行为动作过于发散,其决策结果的成功率偏低的问题。上述两小节为使六足机器人可以基于驾驶员的先验知识快速提升其决策成功率,利用神经网络归约驾驶员决策的先验知识,再把环境和驾驶员因素充分融入到机器人的决策训练,使快速形成其自主决策能力,并不断优化更新先验模型参数,使机器人逐渐具有适应已知地形的驾驶决策智能。

具体的,为了更好地实现上述目标,本节总结了基于先验模型的六足机器人决策训练方法,其特点在于该方法是一种人-机-环耦合交互的决策训练方法。如图 3 所示,首先,利用人与环境交互的结果,建立驾驶员的决策指令和局部环境矩阵建立映射组合,形成训练样本。进一步的,人机间的交互,是机器人利用神经网络归约驾驶员先验决策模型的过程。考虑到本文设计的神经网络隐含结构是基于驾驶员决策逻辑的分层结构,因此需要对每个功能层分别建立损失函数。不同于传统平方差求和的方式建立的损失函数,本文采用交叉熵这一更为先进有效的方法。目的在于采用交叉熵法建立的损失函数可在权重更新阶段更为迅速,也在于采用交叉熵法建立的损失函数属于凸函数,可有效避免陷入局部最优解问题。例如方向判断层,当该层接收到训练样本中的环境信息的输入后,其输出值为 d , 而训练样本输出值 d' , 则其的损失函数可表达为式(10)。为使 3 个判断层和 3 个指令层的损失函数达到最小,本文借助梯度下降法逐层对神经网络参数(权重值与偏置值)反向寻优,设当前神经网络参数权重向量 w^n 与下一次的权重向量 w^{n+1} , 则其通过描述二者之间的关系即为梯度下降参数更新方法,可表达为式(11)。直到输出值与样本值的差小于 1% 的样本值时,机器人得到驾驶员先验模型。最后机器人与环境间以强化学习的方式进行交互,最终得到一个具有较

强化化能力的机器人自主决策模型。

$$L(d, d') = - \sum_{i=1}^a d'_i \ln d_i \quad (10)$$

$$\dot{w}^{n+1} = \dot{w}^n - \frac{\partial L}{\partial w^n} \quad (11)$$

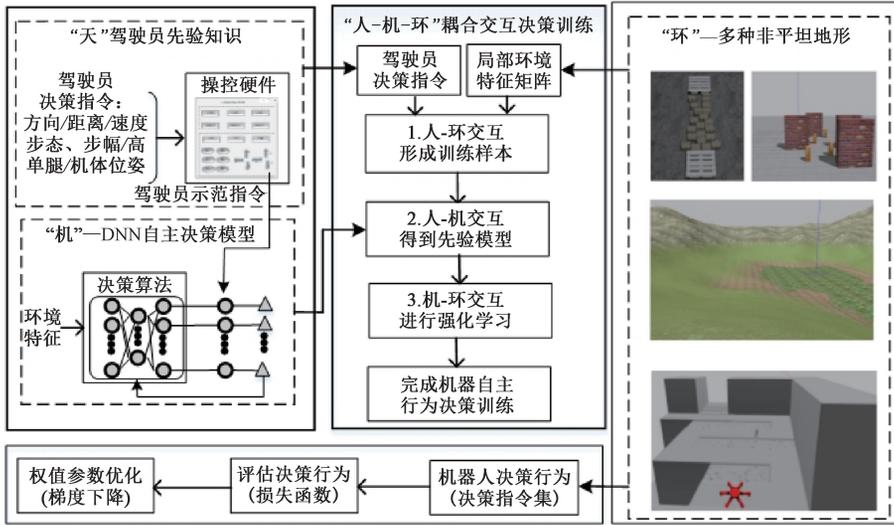


图 3 六足机器人决策训练方法

Fig. 3 Decision training method of the hexapod robot

3 人机协同决策指令生成方法

驾驶员和机器智能的决策指令都具有一定的不确定性,所以依据单方决策指令不能完全保障六足机器人行走运动完全可靠,尤其是无法充分融合人机双方的决策优势。

当双方具有不相矛盾的共同决策指令时,指令将立即发送至位于环境中机器人的运动控制系统中。然而,当人机双方的指令发生冲突时,驾驶员和机器人两个智能主体都倾向于在稳定裕度、能量分配、碰撞次数等不同指标方面得到收益,从而做出利己的决策,此时,人机在总体合作的大背景下形成局部博弈的格局。本节首先利用得到的地形矩阵归纳出局部地形的特征向量,并计算该地形特征下 3 个指标的单机和两极收益,进而利用合作博弈中的 shapley 值法^[17-18]计算出三者的边界收益,最后通过平衡人机对这 3 种指标的收益,生成人机协同决策指令。

具体的, $A_{feature} = (\alpha, h, n)$ 为局部地形的特征向量, α 为地形的平均斜度, h 为地形的平均障碍高度, n 为障碍个数。 $T'_{(3 \times 3)}$ 为单极转换矩阵, $T''_{(3 \times 3)}$ 为双极转换矩阵。 \vec{V}_{single} 为单极收益函数, \vec{V}_{double} 为双极收益函数, $\vec{V}_{tripartite}$ 为三极收益(设为定值为 100), 单极和双极收益的表达式如下:

$$\vec{V}_{single}(V(m), V(c), V(e)) = A_{feature} \cdot T'_{(3 \times 3)} \quad (12)$$

$$\vec{V}_{double}(V(mc), V(me), V(ec)) = A_{feature} \cdot T''_{(3 \times 3)} \quad (13)$$

其中, $V(m)$ 、 $V(c)$ 和 $V(e)$ 分别为稳定裕度收益、碰撞次数受益和能量分配收益。 $V(mc)$ 、 $V(me)$ 和 $V(ec)$ 分别为三者的双极组合收益。

由 shapley 边界贡献函数可求得边界收益:

$$Sh(S, i) = \frac{1}{|Ag|} \sum_{Ag} V(S \cup i) - V(S) \quad (14)$$

其中, $i = m, c, e$; S 为由 m, c, e 成立的组合; $Sh(S, i)$ 代表 i 在 S 组合下用 shapley 值法平衡后的收益值。 Ag 为组合的个数, 由 m, c, e 成立的组合共为 6 个。 $Sh(\{m, c, e\}, m)$ 、 $Sh(\{m, c, e\}, c)$ 和 $Sh(\{m, c, e\}, e)$ 分别为稳定裕度、碰撞次数和能量分配在三极组合中的边界收益, 简记为: $Sh(m)$ 、 $Sh(c)$ 和 $Sh(e)$ 。

根据驾驶员顶层决策的逻辑习惯, 其指令集可表示为 $\{d_h, v_h, \theta_h, \lambda_h, l_h, h_h, \omega_h, \mu_h, \psi_h\}$, 机器智能指令集可表示为 $\{d_r, v_r, \theta_r, \lambda_r, l_r, h_r, \omega_r, \mu_r, \psi_r\}$, 人机协同指令集可表示为 $\{d_{co}, v_{co}, \theta_{co}, \lambda_{co}, l_{co}, h_{co}, \omega_{co}, \mu_{co}, \psi_{co}\}$, 则可利用 3 个边界收益去合成人机协同指令, 如预期距离的人机协同指令 d_{co} 表达为:

$$\begin{cases} d_{co} = P_d \cdot \Delta d \\ \Delta d = d_h - d_r \\ P_d = f(Sh(m), Sh(c), Sh(e)) \end{cases} \quad (15)$$

4 实验

4.1 重载六足机器人人机协同决策半物理仿真系统

重载六足机器人在获得决策智能过程中,若直接采用总重达 0.7 t 实物样机进行实验,其行走安全性很难得到保障。随着数字孪生技术^[19-21]在航空航天等大型制造装备领域被越来越多地应用,虚拟仿真系统与实物样机一样,都分别承担着不可被替代的功能。尤其是当本文聚焦于讨论提升机器人的顶层决策智能的方法,若直接使用样机开展验证实验,很难避免将底层控制误差干扰顶层决策指令的实施效果。然而,仅依靠传统虚拟仿真软件,又难以满足驾驶员对于操控决策的操控临场感需求。因此,本文将虚拟的场景仿真和真实的操控硬件相结合,首次面向大尺度重载六足机器人的人机协同问题,搭建了人机协同驾驶决策半物理仿真系统。如图 4 所示,在仿真计算机内的 ROS 系统中利用 Gazebo 软件对重载六足机器人进行建立虚拟仿真模型,不仅保证其底层运动控制的准确性,同时也实现了对于虚拟数字地形环境的仿真还原。此外,将操控推杆、操控摇杆以及可触摸的人机交互界面接入到仿真控制程序,深度还原驾驶员在驾驶舱内进行决策的真实操控的情景。下一阶段,团队在后续迁移部署的方式上,将充分汲取瑞士联邦理工大学四足机器人 ANYmal,将强化训练结果由仿真系统向实物机器人迁移部署时的技术模式^[22],通过采用两次“仿真训练-实物验证”闭环部署方式将基于强化学习的决策框架逐步部署到重载实物样机中。即借助小型六足机器人及其仿真训练环境完成第一轮算法验证,再借助所搭建的人机协同驾驶决策半物理仿真系统和大尺度六足机器人样机完成第二轮算法验证。



图 4 重载六足机器人人机协同驾驶决策半物理仿真系统

Fig. 4 Half-physical simulation system for the heavy-duty hexapod robot

4.2 野外复杂地形下的驾驶决策训练与人机驾驶决策实验

为了能充分学习驾驶员在野外山地环境下的驾驶决策经验,本文将障碍地形、斜坡地形、沟壑地形以及台阶地形融入到山地环境的大场景中,并利用本文所提出的方法在上述常见的 4 种非规则地形下进行训练,如图 5 所示。

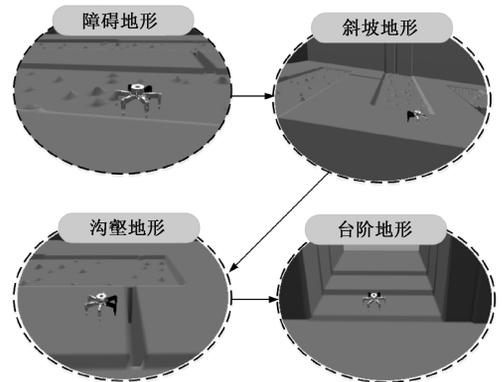


图 5 虚拟野外山地环境

Fig. 5 Virtual wild mountain environment

在此基础上,开展有人决策驾驶、机器智能自主决策及人机协同决策在相同地形下的对比实验,并在稳定裕度、碰撞次数和能量消耗 3 个维度对驾驶员、机器智能及人机协同智能的驾驶决策结果进行评价。具体的,稳定裕度是指机身重心在水平面内的垂直投影到支撑模式边界的最短距离,即静态稳定裕度;碰撞次数是指机器人的胫节与环境中障碍物的累计接触次数;能量消耗是根据记录各关节的角位移,并结合关节电机的平均转矩进行得出估计值。

4.3 实验结果与分析

如图 6~9 和表 1 所示,驾驶员决策、机器智能自主决策以及人机协同决策指令结果在稳定性上的评测。可以看出,经过本文所提出方法进行训练后,六足机器人利用得到的驾驶员决策的神经网络模型,在稳定裕度评测维度上,自主决策的效果可基本接近有人驾驶决策的效果。但在障碍、沟壑和台阶地形下,自主决策的稳定裕度略低于有人驾驶决策,而人机协同决策指令效果介于二者之间。这是由于机器人没有像驾驶员一样,在非规则地形下更趋于做出谨慎的决策与操作,而通过人机协同决策可有效弥补机器人在这方面的决策缺陷。

如图 10~13 所示,驾驶员决策、机器智能自主决策以及人机协同决策指令结果在撞次数上的评测。由于在仿真系统中,将机器人胫节与环境障碍的接触视为碰撞,并且为了降低碰撞的未检率,提高测评的安全系数,将系统的检测频率设定为 10 Hz,因此所得到的检测值略高

表 1 3 种决策方式的稳定裕度

Table 1 Stability margin of three decision-making methods

决策智能体	障碍地形	斜坡地形	沟壑地形	台阶地形
驾驶员	0.442	0.497	0.462	0.455
机器智能	0.429	0.492	0.441	0.439
人机协同	0.439	0.495	0.449	0.442

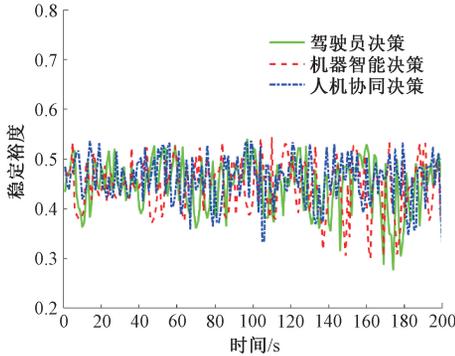


图 6 障碍地形下 3 种决策方式的稳定裕度

Fig. 6 Stability margin of the three decision-making methods in the obstacle terrain

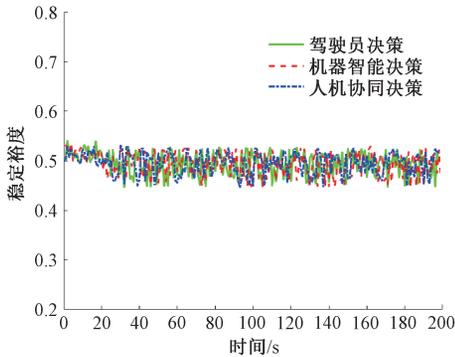


图 7 斜坡地形下 3 种决策方式的稳定裕度

Fig. 7 Stability margin of the three decision-making methods in the slope terrain

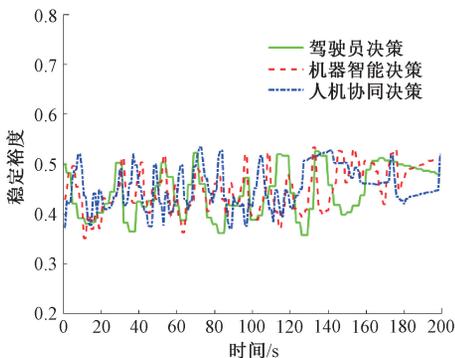


图 8 沟壑地形下 3 种决策方式的稳定裕度

Fig. 8 Stability margin of the three decision-making methods in the ditch terrain

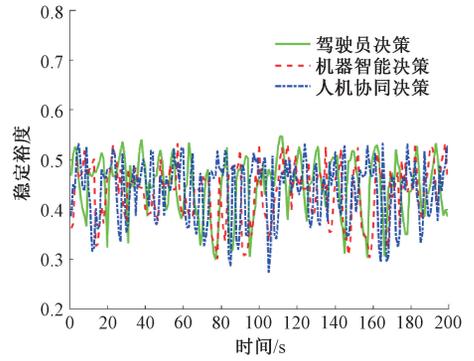


图 9 台阶地形下 3 种决策方式的稳定裕度

Fig. 9 Stability margin of the three decision-making methods in the stairs terrain

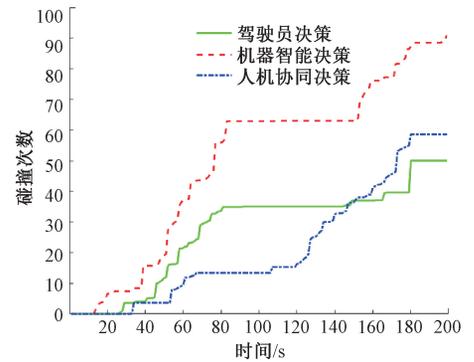


图 10 障碍地形下 3 种决策方式的碰撞次数

Fig. 10 Collision number of the three decision-making methods under the obstacle terrain

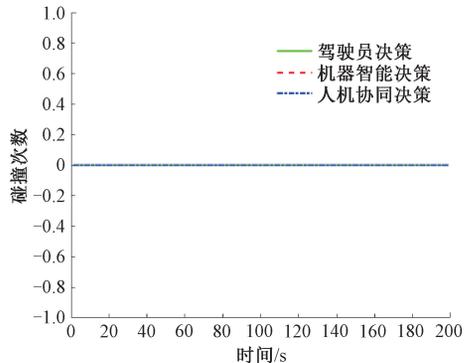


图 11 斜坡地形下 3 种决策方式的碰撞次数

Fig. 11 Collision number of the three decision-making methods in the slope terrain

于真实碰撞次数。可以看出,经过本文所提出方法进行训练后,六足机器人在斜坡地形时,与有人驾驶决策和人机协同决策的效果相同,均未发生与环境的碰撞;

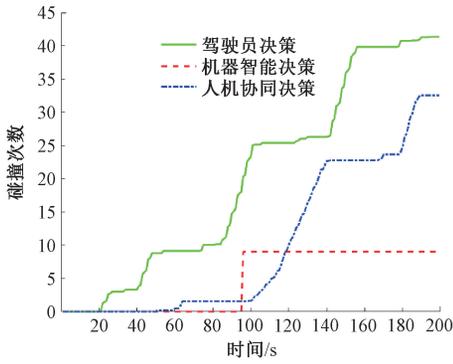


图 12 沟壑地形下 3 种决策方式的碰撞次数

Fig. 12 Collision number of the three decision-making methods in the ditch terrain

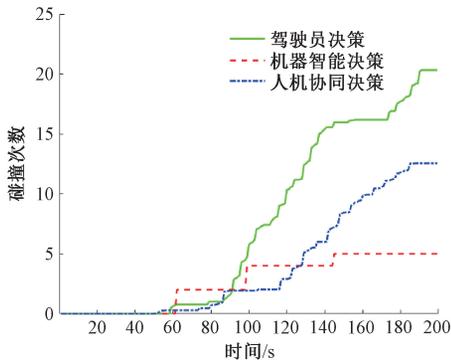


图 13 台阶地形下 3 种决策方式的碰撞次数

Fig. 13 Collision number of the three decision-making methods in the ditch terrain

在沟壑和台阶这种相对规则的非平坦地形下,机器人自主决策在碰撞次数上甚至略优于有人驾驶的效果,人机协同决策的效果介于二者之间。这是由于机器人在面对规则障碍时的重复定位精度高于驾驶员。但在障碍地形下,由于障碍物的分布具有一定的不均匀分布特性,自主决策的碰撞次数略高于人驾驶决策。由此说明,对于随机分布的非规则障碍地形,而通过人机协同决策可有效弥补机器人或驾驶员在这方面的决策缺陷。

如图 14~17 所示,驾驶员决策、机器智能自主决策以及人机协同决策指令结果在能量消耗上的评测。可以看出,经过本文所提出方法进行训练后,六足机器人利用得到的驾驶员决策的神经网络模型,在能量消耗评测维度上,其自主决策的效果基本接近有人驾驶决策的效果。尤其是在斜坡、沟壑地形下,三者的稳能力消耗接近。甚至在台阶这种规则障碍下,驾驶员决策的能量消耗略高于机器人。然而在障碍地形下,自主决策的能量消耗依

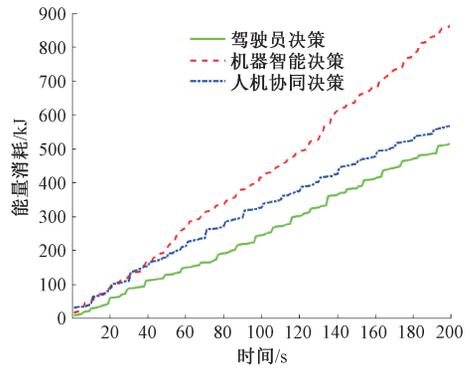


图 14 障碍地形下 3 种决策方式的能量消耗

Fig. 14 Energy consumption of the three decision-making methods in the obstacle terrain

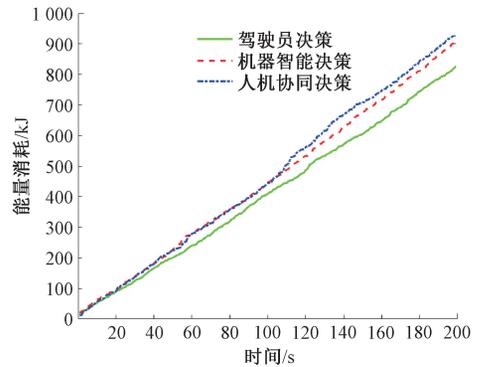


图 15 斜坡地形下 3 种决策方式的能量消耗

Fig. 15 Energy consumption of the three decision-making methods in the obstacle terrain

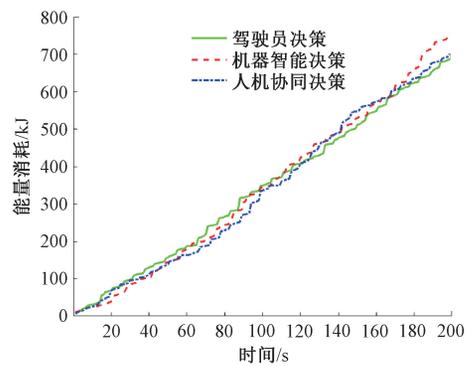


图 16 沟壑地形下 3 种决策方式的能量消耗

Fig. 16 Energy consumption of the three decision-making methods in the ditch terrain

然略高于有人驾驶决策,而人机协同决策可有效弥补机器人这方面的决策缺陷。

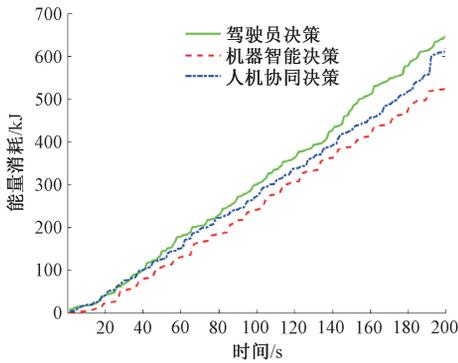


图 17 台阶地形下 3 种决策方式的能量消耗

Fig. 17 Energy consumption of the three decision-making methods in the stairs terrain

5 结 论

本文聚焦于利用驾驶员先验模型提高重载六足机器人在复杂环境下的自主决策能力以及人机协同决策指令的融合度。在人-机-环三者的耦合交互的过程中,驾驶员与环境交互形成训练样本;机器人与驾驶员进行交互学习,形成驾驶员决策先验模型;机器人与环境进行交互,更新先验模型,并提高自主决策的泛化能力。实验结果表明机器智能通过学习驾驶员先验模型和自主训练,在训练地形下其决策效果可接近驾驶员决策效果,同时人机协同决策指令可有效弥补单智能体决策指令的缺陷。

参考文献

- [1] HALME A, VAINIO M. Forestry robotics-why, what and when[M]. London: Springer, 1998.
- [2] WILCOX B H. ATHLETE: A cargo-handling vehicle for solar system exploration[C]. IEEE Aerospace Conference, IEEE, 2011: 5-12.
- [3] PICARDI G, CHELLAPURATH M, IACOPONI S, et al. Bioinspired underwater legged robot for seabed exploration with low environmental disturbance [J]. Science Robotics, 2020, 5(42):1012.
- [4] LI J, YOU B, DING L, et al. Dual-master/single-slave haptic teleoperation system for semiautonomous bilateral control of hexapod robot subject to deformable rough terrain[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 21(4): 1-15.
- [5] LI J, YOU B, DING L, et al. A novel bilateral haptic teleoperation approach for hexapod robot walking and manipulating with legs [J]. Robotics and Autonomous Systems, 2018, 108:1-12.
- [6] HWANGBO J, LEE J, DOSOVITSKIY A, et al. Learning agile and dynamic motor skills for legged robots[J]. Science Robotics, 2019, 4(26):5872.
- [7] KUMAR A, FU Z, PATHAK D, et al. RMA rapid motor adaptation for legged robots[C]. Robotics-Science and Systems, 2021.
- [8] CULLY A, CLUNE J, TARAPORE D, et al. Robots that can adapt like animals[J]. Nature, 2015, 521(7553): 503-507.
- [9] CHEN Y, WU Y, QIN Y, et al. Recognizing nested named entity based on the neural network boundary assembling model[J]. IEEE Intelligent Systems, 2019, 35(1):74-81.
- [10] REGE A, OBRADOVIC Z, ASADI N, et al. Predicting adversarial cyber intrusion stages using autoregressive neural networks [J]. IEEE Intelligent Systems, 2018, 32(2):29-39.
- [11] TAYLOR A, LEBLANC S P, JAPKOWICZ N. Probing the limits of anomaly detectors for automobiles with a cyberattack framework [J]. IEEE Intelligent Systems, 2018,32(2):54-62.
- [12] SANCHEZ R G. Hydraulically actuated hexapod robots: Design, implementation and control [J]. Computing Reviews, 2014, 55(9):545-545.
- [13] KURISU M. A study on teleoperation system for a hexapod robot — Development of a prototype platform [C]. International Conference on Mechatronics & Automation, IEEE, 2011.
- [14] YHOKOI K, KAJITA S. Gait planning for energy efficiency in walking machines[J]. IEEE Transactions on Robotics, 2007, 23(4): 763-775.
- [15] 尤波,曲伟健,李佳钰. 面向双操作者的六足机器人共享遥操作[J]. 控制与决策,2022,37(11):2769-2778. YOU B, QU W J, LI J Y. Research on shared teleoperation of hexapod robot for dual operators [J]. Control and Decision,2022,37(11):2769-2778.
- [16] 尤波,陈翰南,李佳钰,等. 一种基于变权重的六足机器人共享遥操作控制[J]. 仪器仪表学报,2019,

40(8):239-250.

YOU B, CHEN H N, LI J Y, et al. Shared teleoperation control of hexapod robot based on variable weight [J]. Chinese Journal of Scientific Instrument, 2019, 40(8): 239-250.

[17] LIU H, ZHANG Q. Cooperative games with additive multiple attributes [J]. Journal of Intelligent and Fuzzy Systems, 2021, 41(1):1135-1150.

[18] YE Y F, LI D F. A direct approach to compute triangular fuzzy banzhaf values of cooperative games with coalitions' values represented by triangular fuzzy numbers [J]. IEEE Transactions on Fuzzy Systems, 2021, 29(6): 1567-1575.

[19] 刘大同, 郭凯, 王本宽. 数字孪生技术综述与展望 [J]. 仪器仪表学报, 2018, 39(11):10-20.

LIU D T, GUO K, WANG B K. Summary and perspective survey on digital twin technology [J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 10-20.

[20] WU J, YANG Y, CHENG X, et al. The development of digital twin technology review [C]. Chinese Automation Congress (CAC), 2020.

[21] REICHE L T, GUNDLACH C S. The digital twin of a system: A structure for networks of digital twins [C]. 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2021: 1-8.

[22] LEE J, HWANGBO J, WELLHAUSEN L, et al. Learning quadrupedal locomotion over challenging terrain [J]. Science Robotics, 2020, 5(47):5986.

作者简介



陈潇磊, 2015年于燕山大学获得学士学位, 现为哈尔滨理工大学博士研究生, 主要研究方向为足式机器人智能控制与人机协同技术。

E-mail: cxl_920101@163.com。

Chen Xiaolei received his B.Sc. degree from Yanshan University in 2015. He is currently a Ph.D. candidate at Harbin University of Science and Technology. His research interests include hexapod robot intelligent control, and human-robot collaboration.



尤波 (通信作者), 1995年于哈尔滨工业大学获得博士学位, 现为哈尔滨理工大学教授, 主要研究方向为智能机器人与机电控制。

E-mail: youbo@hrbust.edu.cn

You Bo (Corresponding author) received his Ph.D. degree from Harbin Institution of Technology in 1995. He is currently a professor at Harbin University of Science and Technology. His research interests include intelligent robot and mechatronics system.



丁亮, 2009年于哈尔滨工业大学获得博士学位, 现为哈尔滨工业大学教授, 主要研究方向为地面力学与智能移动机器人。

E-mail: liangding@hit.edu.cn

Ding Liang received his Ph.D. degree from Harbin Institution of Technology in 2009. He is currently a professor at Harbin Institution of Technology. His main research interests include ground mechanics and intelligent mobile robot.