



具有学习和推理能力的空间无人系统智能架构

黄煌^{1,2}, 李谋^{1,2*}, 刘磊^{1,2}, 汤亮^{1,2}, 刘昊^{1,2}, 谢心如^{1,2}, 刘乃龙^{1,2},
魏春岭^{1,2}, 邢琰^{1,2}, 姜甜甜^{1,2}, 胡海东^{1,2}, 常亚菲^{1,2}, 胡勇^{1,2}, 杨孟飞³

1. 北京控制工程研究所, 北京 100190

2. 空间智能控制技术重点实验室, 北京 100094

3. 中国空间技术研究院, 北京 100094

* 通信作者. E-mail: limou2333@163.com

收稿日期: 2022-01-21; 修回日期: 2022-04-01; 接受日期: 2022-05-02; 网络出版日期: 2022-11-10

科技部新一代人工智能重大项目 (批准号: 2018AAA0102700) 资助

摘要 以卫星、飞船、深空探测器为代表的航天器是一类典型的无人系统, 经历了从自动化到自主化的不断发展. 为提升航天器对未知空间环境和复杂空间任务的适应能力和智能自主水平, 以空间无人系统为研究对象设计了具有学习和推理能力的无人系统智能架构, 给出架构的组成及功能; 接着分析该架构中学习和推理能力的运作机制, 并针对架构所需的动作库和知识库的构建、更新与扩展方法等关键技术提出解决方案; 最后通过地外星球巡视采样任务示例, 具体说明在该智能架构下的新知识生成、基于知识的推理和自主执行任务的过程, 以及过程中知识库、环境、任务和动作库之间的交互关系.

关键词 学习和推理能力, 空间无人系统, 智能架构, 动作库, 知识库

1 引言

随着人工智能技术的迅速发展, 智能无人系统在无人汽车、无人飞机等领域已经有了广泛应用, 典型代表有特斯拉实现了较高自主等级自动驾驶的道路测试^[1], Alpha 空战系统中 AI 驱动的无人机战胜了有经验的飞行员, 甚至自主习得“大方位角射击”这样一种全新的攻击样式^[2]. 以卫星、飞船、深空探测器等为代表的航天器作为一类典型的无人系统, 工作于空间环境或地外星球环境, 具有严苛环境未知、先验知识欠缺、通信条件恶劣等特点, 是智能技术应用的重要领域. 杨嘉墀^[3]在 1995 年发表的《中国空间计划中智能自主控制技术的发展》一文中, 论述了智能自主之于航天器的重要意义. 为应对地外天体探测、在轨维修维护、非合作目标处理等日益复杂的空间任务需求, 提升航天器在未

引用格式: 黄煌, 李谋, 刘磊, 等. 具有学习和推理能力的空间无人系统智能架构. 中国科学: 信息科学, 2022, 52: 2093-2105, doi: 10.1360/SSI-2022-0036
Huang H, Li M, Liu L, et al. Intelligent architecture of space unmanned system with learning and reasoning capabilities (in Chinese). Sci Sin Inform, 2022, 52: 2093-2105, doi: 10.1360/SSI-2022-0036

知空间环境和复杂空间任务下的适应能力, 迫切需要构建一套会学习、能推理的无人系统智能架构以及相应的智能实现框架和生成机制。

2016 年我国发布第四部航天白皮书^[4], 明确提出建设航天强国的战略目标; 2017 年我国发布《新一代人工智能发展规划》^[5], 计划到 2030 年人工智能理论、技术与应用总体达到世界领先水平。作为战略前沿技术, 人工智能技术是推动航天事业进一步发展的重要手段, 特别是随着硬件水平的提高和近年来人工智能理论、方法与技术的飞速发展, 为人工智能技术在航天领域的应用提供了良好的软硬件基础。随着智能感知、自主决策、重构升级等能力成为未来航天器的基本能力, 发展新一代智能无人系统, 深度融合人工智能技术和航天技术, 将加快航天强国建设并促进智能无人系统的发展。

通常认为智能无人系统是由无人平台及若干辅助部分组成, 具有感知、交互和学习能力, 并且能够基于知识进行自主推断、自主决策, 从而达成目标的有机整体^[6]。智能无人系统的应用主要集中在机器人、无人车、无人机、无人船、无人潜航器等领域, 目前已应用的无人系统多为自动化或自主系统, 如美国波士顿动力公司 ATLAS 仿人类机器人^[7]、谷歌公司的 Google Driverless Car 全自动驾驶汽车^[8]、中国航天科技集团“彩虹”系列无人机^[9]、中国科学院沈阳自动化研究所“海斗一号”全海深自主遥控潜水器^[10]等。这些系统主要工作在相对确定的环境下, 能够自主完成目标识别、避障、路径规划等任务, 但在与环境交互的场景理解、推理决策和自主学习等方面还有待进一步提升。

在航天领域的人工智能应用研究始于 20 世纪 70 年代, 1977 年, 美国国家航空航天局 (National Aeronautics and Space Administration, NASA) 机器智能与机器人学的研究小组受命研究人工智能对整个航天领域的影响^[11], 随后主导并开展了一系列的应用研究。目前, 人工智能技术在航天领域的应用主要包括运载火箭、卫星、飞船、空间探测、宇航系统设计与制造等方面, 其中卫星和空间探测是智能无人系统应用的重点领域, 用于执行星上自主感知、数据处理、规划控制、故障诊断、多星协同等任务。如 NASA 于 1998 年发射的深空 1 号卫星^[12], 验证了自主导航和自主远程代理等多项自主技术, 实现了一定程度的自主规划、故障检测诊断等功能; 2003 年发射的勇气号火星探测器^[13], 在轨运行 7 个月后安全登陆火星, 并开展了长达 7 年的探测任务; 2020 年发射的毅力号火星探测器^[14], 其装备的 AutoNav 系统具有一定的“思考”能力, 可实现自动驾驶、移动中决策和主动规避潜在危险障碍等功能。

我国的航天器智能自主控制的研究从 20 世纪 90 年代开始^[15], 在自主导航、卫星姿态和轨道控制、故障诊断等方面取得了多项研究成果^[16]。例如吉林 1 号光学遥感卫星^[17], 具备在轨智能信息处理能力, 能够在轨快速识别、定位关键目标; 在探月工程中实现了月球探测器的着陆、起飞、上升阶段的自主控制, 玉兔号月球车自主路径规划^[18]; 2020 年发射的天问一号火星探测器^[19], 自主完成了火星着陆, 祝融号火星车可以自主规划行驶路线。2019 年北京控制工程研究所从运动控制、感知认知、决策规划等维度, 结合新时期人工智能技术发展的特点及空间任务的新需求, 提出了空间智能控制系统的 5 级分级^[20], 分析认为, 当前我国已掌握自主交会对接、月面自主软着陆技术, 在轨实现了卫星典型特征的识别、空间站与神舟飞船变结构体的控制, 还在完善等级 3, 同时也在向等级 4 发展。相比之下, 美国已具备星上目标识别、自主机动能力, 智能自主水平整体上处于等级 3, 当前正在向等级 4 发展。

在无人系统架构设计方面, Chiang 等^[21]基于 3-T 架构^[22]提出一种具有分层结构的自适应智能架构, 主要用于 NASA 的移动机器人, 但也适用于一般智能控制; Hassani 等^[23]为观测卫星的星上自主运行设计了一种包括决策层、执行层和功能层的智能架构, 将低级物理行为与概念决策能力相结合; Narayan 等^[24]针对无人机系统提出了一种新的多层智能控制体系结构, 其中最高层提供深思熟虑的推理 (包括态势感知和任务规划), 中间层实现导航 (如路径规划和操纵生成), 功能控制层提供反应

功能, 实现强制着陆和避免碰撞; Qureshi 等^[25] 针对非合作目标交会对接任务, 将视觉感知与高级推理结合提出一种慎思/反应混合计算框架, 并在实验室环境中实现机械臂对自由飞行卫星的自动捕获. 施等^[26] 提出建立多层分簇的协同框架帮助无人系统面对变化时快速反应. 这些无人系统架构, 借鉴了人类的反应-慎思行为模式, 描述了智能无人系统的设计理念, 但是在知识的生成、自主学习的实现机制等方面, 还有待进一步深入的研究.

综上, 本文围绕复杂空间任务对航天器智能自主能力的迫切需求, 从空间智能无人系统的基本组成、逻辑关系、信息流和学习机制等方面出发, 提出具有学习和推理能力的空间无人系统智能架构 (以下简称“智能架构”), 并对智能架构中的关键技术提出可行的解决方案, 以期为空间无人系统智能自主能力的提升提供一种实现框架和生成机制.

2 智能架构组成和定义

目前, 部署在航天器上的智能无人系统或智能算法多为功能性的, 通常只能实现某些任务中的某类特定功能, 如目标识别、路径规划、故障诊断等. 由于没有进行系统性的设计和考虑, 现有的空间无人系统在知识的有效利用和生成等方面的能力欠佳, 对未知空间环境的适应能力有限, 在执行复杂空间任务时对地面人员的依赖程度仍然较高, 还难以达到任务级的智能自主, 如毅力号火星探测器已经具有较强的自主导航能力, 但探测过程中何时采集样品、何时巡视探测, 仍然需要地面站进行规划. 本文提出的智能架构从基本原理和各模块的运作机制方面开展研究, 拟解决知识如何作用到智能体并影响智能体的行为和决策等问题, 使无人系统具有任务级的学习和推理能力.

一般涉及智能体与环境交互的学习过程均可归结为马尔可夫决策过程 (Markov decision process, MDP)^[27], 其核心在于将目标导向的行为的学习问题抽象为智能体与环境之间交互的行动、状态和评价 3 个信号, 具有较强的普适性和有效性. MDP 的一个主要性质是其当前状态和评价中每个可能出现的概率仅与前一个状态和前一个动作有关, 而与更早的状态和动作完全无关. 但对于相同的任务, 状态和动作表征方式及评价机制的选取会显著影响其学习过程, 目前还没有选取方法的一般原则, 主要依靠设计人员的经验和技巧.

具有学习和推理能力的空间无人系统智能架构如图 1 所示, 智能架构由智能体、环境和任务等部分组成, 智能体是智能架构的核心, 智能体主要包含动作库、知识库、感知机、学习推理机和评估模块等部分, 动作库和知识库是实现学习和推理能力的基础. 智能体加载有特定的知识并能够执行一定的动作. 智能体利用感知机从传感器数据中提取环境特征, 学习推理机根据环境特征在知识库中搜索相关知识并生成要执行的动作, 评估模块评估整个任务执行过程并给出任务执行的优劣评价指标, 学习推理机再根据评价指标更新、扩展动作库和知识库并自更新, 最终使评价指标向期望的方向变化.

综上, 智能架构中智能体利用现有的知识库理解环境和任务的状态和变化, 进而推理出应执行的动作, 并在与环境 and 任务的交互过程中实现自我更新和能力提升, 达到适应不同的环境和任务的目的. 本文给出智能架构的定义如下.

定义 1 具有学习和推理能力的无人系统智能架构, 由智能体、环境和任务 3 部分组成, 其中智能体包括知识库、动作库、感知机、学习推理机和评估模块, 智能体具有在线学习能力, 能够在与环境 and 任务交互过程中生成新知识和新动作.

在定义 1 的智能架构中, 动作库包含智能体所能够执行的各种动作, 是智能体与环境 and 任务的交互手段. 动作库中的动作分为元动作和子动作, 描述了智能体的具体结构、功能和行为. 其中, (1) 元动作是动作中相互独立的基本单元, 表示的是智能体的物理特征, 一个智能体的全部元动作构成这个

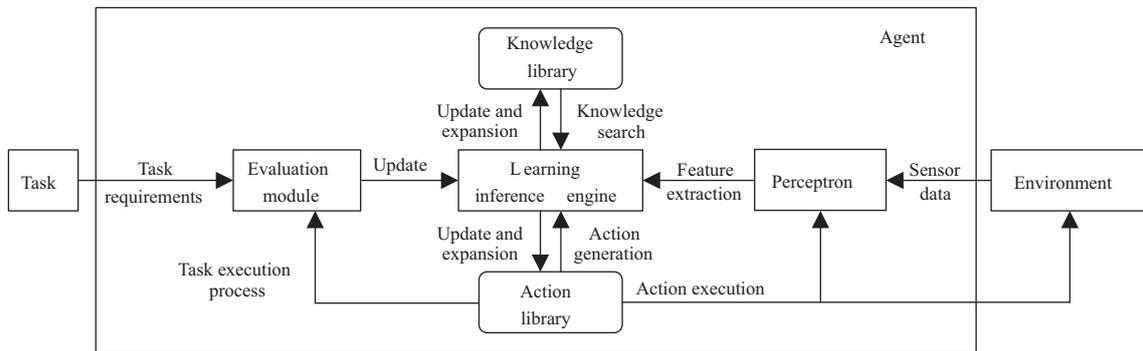


图 1 具有学习和推理能力的空间无人系统智能架构

Figure 1 Intelligent architecture of space unmanned system with learning and reasoning capabilities

智能体的元动作集, 表征一个智能体的元动作集应是正交且完备的, 具有相同元动作集的智能体是同一种智能体, 同一种智能体的功能相同, 但不同个体间可以存在强弱差异; (2) 子动作是有机组合的元动作序列, 子动作集的构成与具体的环境和任务相关, 表示的是智能体对环境、任务及自身的理解和适应, 智能体应该具有根据环境和任务的变化更新、扩展子动作集的能力。

在定义 1 的智能架构中, 知识库包含智能体的所有知识, 是对物理环境特征的描述 (这里物理环境泛指包括智能体的物理世界, 即世界模型), 知识分为元知识和子知识, 同一种智能体可以加载不同的知识库, 形成特化的智能体. 其中, (1) 元知识是对世界模型中抽象特征的描述, 表示的是“是什么”的含义, 元知识间相互独立, 由于世界模型的复杂性和未知性, 无法建立正交且完备的元知识集, 但可以根据具体任务建立匹配的元知识集, 用于描述局部的世界模型, 同时元知识集应该具有可更新和扩展性; (2) 子知识是元知识的连接关系和组织形式, 表示的是因果关系、基本规律、求解步骤等含义, 也体现智能体对任务、环境及自身的理解和适应, 同时子知识集也应具有可更新和扩展性。

在动作库和知识库的支撑下, 智能体将具有学习和推理两个基本能力. 本文给出智能架构中学习能力和推理能力的定义如下。

定义 2 对于定义 1 给出的智能架构, 其学习能力是指能够根据已有的数据建立任务、环境、知识和动作的映射关系, 推理能力是指能够通过和环境的交互更新、扩展动作库和知识库 (尤其是元知识集), 从而提高智能体对未知环境和复杂任务的适应能力。

需要特别指出的是, 虽然该智能架构提出的初衷是提升航天器对未知空间环境和复杂空间任务的适应能力和智能自主水平, 但该智能架构仍可适用于除空间无人系统之外的机器人、无人车、无人机、无人船、无人潜航器等领域, 是一个具有通用性的智能架构。

3 智能架构的运作机制和关键技术

根据智能架构的组织形式, 为实现智能架构的学习和推理能力, 需要明确智能体学习和推理机制, 解决动作库和知识库的构建与更新、扩展方法等关键技术。

3.1 学习推理机的运作机制

在智能体学习和推理机制方面, 建立在数理统计上的机器学习和建立在符号表达上的逻辑推理是两大主流技术路线, 前者重学习, 依赖大量带标签的数据; 后者重推理, 善于利用规则知识. 智能架构

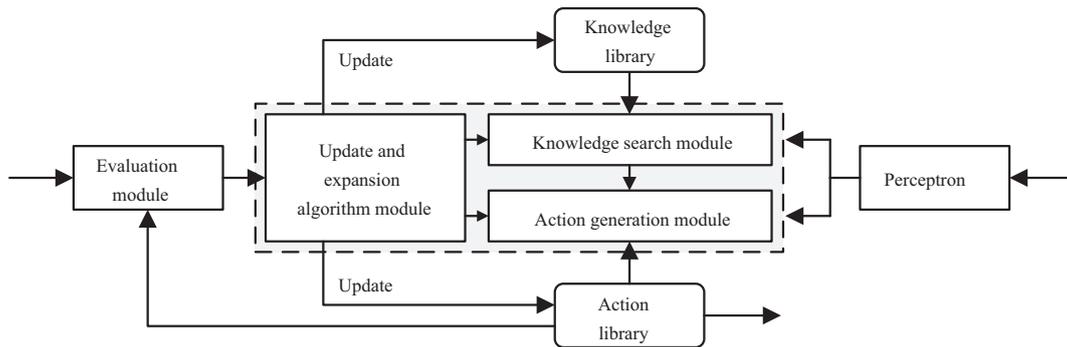


图 2 智能体内信息流

Figure 2 Information flow of the agent

的学习和推理机制应该同时结合两者的优点。

智能架构中学习推理机应包括知识搜索模块、动作生成模块和更新扩展算法模块,知识搜索模块利用提取的特征激活知识库中的相关元知识和子知识节点构建连接关系,实现知识搜索和特征匹配的功能;动作生成模块利用提取的特征和相关知识生成要执行的动作(标准元动作控制向量块,详见 3.2 小节),实现知识解释的功能;更新扩展算法模块根据评估结果更新、扩展知识库和动作库,并更新知识搜索模块和动作生成模块。智能体内信息流及学习推理机的内部组织关系如图 2 所示,在智能体的推理决策过程中存在两个回路,一是不经过知识搜索模块的反应式回路,主要用于处理简单事件或敏感事件,具有较高的优先级,通过针对特定任务或场景设立固定规则或安全机制,使智能体可以利用感知机提取的特征直接生成设计好的动作,以提高系统的容错率和可靠性;二是经过知识搜索模块的慎思式回路,主要用于处理复杂事件,通过基于知识的逻辑思考和推理生成动作,以提高系统的适应性和灵活性。这样一种类人的反应式和慎思式双回路设计,融合了传统方法在可靠性和学习方法在适应性上的优势,保证了系统安全可靠运行,同时赋予系统在动态不确定环境中主动探索、自主学习和持续提升的能力。

根据图 2 所示信息流,智能体通过在大范围变化环境下反复执行任务,并不断更新智能体的动作库、知识库和学习推理机,使智能体充分适应不同的环境条件、提高任务执行效率和成功率。

3.2 动作库构建、更新与扩展

在动作库的构建、更新与扩展方法方面,针对无人机、仿生机器人等对象,有学者提出并建立了简易的动作库^[28~30],并产生了元动作、子动作的思想,但并没有给出准确定义。为了描述动作库的更新过程,本文将子动作分为 A、B 两类,A 类是根据任务需求和先验知识人为设计的,B 类是任务执行和环境交互过程中智能体自主产生的。

元动作与子动作的组织关系如图 3 所示,子动作由多个串行执行的标准元动作控制向量块构成,每个块内部又包含多个并行执行的标准元动作控制向量。元动作控制向量作为一条基本的元动作控制指令,由交互部件 ID、开始时间、持续时间、控制信号等部分组成。根据交互部件类别的不同,控制信号可以有不同的形式,如运动执行机构的运动速度、测量敏感器的开关、信号输出设备的信号地址等。对于某个元动作,从不同角度描述的标准元动作控制向量可以是不同的。

智能体在某一时刻输出的动作控制指令并非完整的子动作或者单一的标准元动作控制向量,而是标准元动作控制向量块,经由解析器对块进行解析,得到与交互部件对应的控制信号,再并行分发给交

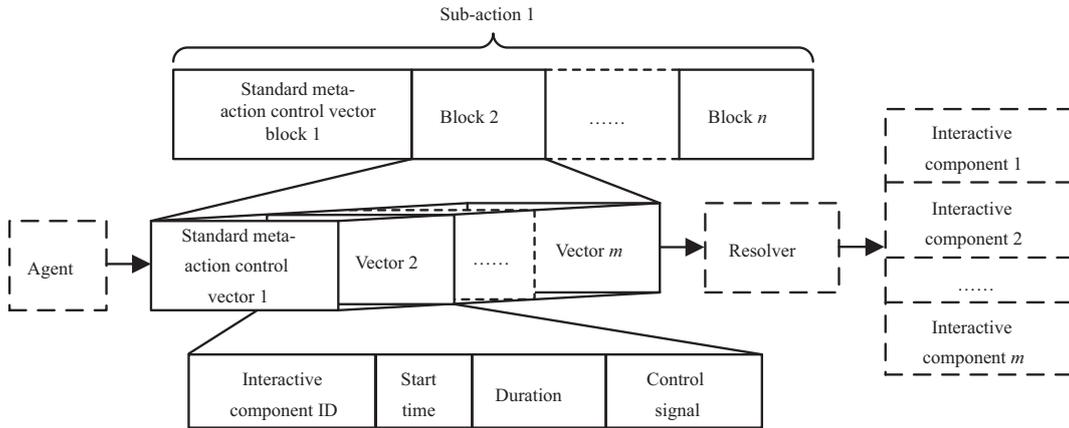


图 3 元动作与子动作的组织关系

Figure 3 Organizational relationship between meta-actions and the sub-action

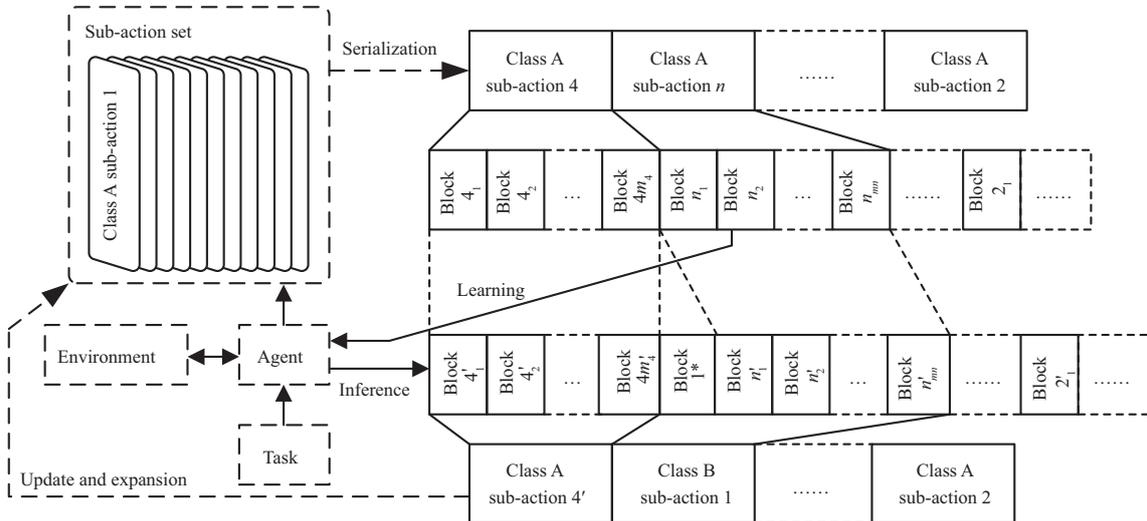


图 4 子动作集更新、扩展过程

Figure 4 Update and expansion process of the sub-action set

互部件执行. 将智能体执行某个具体动作的方式类比为人的动作执行过程, 即在一段时间内持续地给感官系统 (敏感器) 和肌肉系统 (运动执行机构、信号输出设备) 等并行地输出和执行分批块的动作控制指令.

依据图 3 所示的组织关系, 针对具体的任务需求构建子动作集, 并将任务的执行过程分解为子动作执行序列, 令智能体在设定环境条件下学习人为规划的任务执行过程. 由于实际的任务环境和设定条件存在差异, 智能体在实际执行任务过程中产生的块序列与规划的结果会存在一定的差异. 通过对比差异程度和评估任务完成情况, 提取与子动作集中 A 类子动作具有一定相似度的块序列对 A 类子动作进行更新; 提取与子动作集中 A 类子动作具有明显差异且重复出现的块序列扩展为 B 类子动作. 图 4 表示了子动作集的更新、扩展过程, 通过反复迭代学习形成了适应环境和任务的稳定的子动作集, 动作库更新、扩展算法的伪代码如算法 1 所示.

算法 1 Update and expansion algorithm of the sub-action

Initialization: Meta-action set \mathbf{A}_m , sub-action set \mathbf{A}_s , task execution process T , task completion indicator p , similarity ε , threshold γ_A , $k_a = 0$, $k_b = 0$;

Task execution:

- 1: Generate standard meta-action control vector block sequences T' , execute T' to the end of the task, and calculate task completion indicator p' ;
- 2: **if** $p' \geq p$ **then**
- 3: Compare T and T' ;
- 4: **if** there is a sequence \mathbf{a} in T' , and the similarity between \mathbf{a} and the action τ in \mathbf{A}_s is greater than ε **then**
- 5: **if** $k_a > \gamma_A$ **then**
- 6: $\tau \leftarrow \mathbf{a}$;
- 7: **else**
- 8: $k_a \leftarrow k_a + 1$;
- 9: **end if**
- 10: **end if**
- 11: **if** there is a sequence \mathbf{b} in T' , and the similarity between \mathbf{b} and all actions in \mathbf{A}_s is less than ε **then**
- 12: **if** $k_b > \gamma_A$ **then**
- 13: Add \mathbf{b} to \mathbf{A}_s ;
- 14: **else**
- 15: $k_b \leftarrow k_b + 1$;
- 16: **end if**
- 17: **end if**
- 18: **end if**

3.3 知识库构建、更新与扩展

智能体中的知识库与动作库的组织关系如图 5 所示, 在智能体执行具体的任务时, 通过与环境的交互获取环境信息, 并激活知识库中的相关的元知识节点和节点间的连接, 接着根据这些激活的节点和连接从动作库中选取要执行的动作, 装填控制信息后封装成元动作控制向量块输出给交互部件执行。

由于难以面向所有的任务建立正交且完备的元知识集, 区别于元动作集的唯一性和不可扩展性, 在任务执行和环境交互过程中, 智能体的元知识和子知识都应该具有可更新和扩展性. 知识库的更新扩展过程示意如图 6 所示, 智能体和环境交互后提取出相应的特征, 这些特征与知识库中的节点存在一定的对应关系, 对应关系紧密的节点将被激活; 智能体提取的特征中还可能会有和现有的元知识节点都不匹配的特征, 当这些特征重复出现时, 则需要生成与该特征匹配的新的元知识节点, 并与相关的旧的元知识节点建立连接关系; 此外在和环境交互后, 知识库中节点间的连接关系将被重新计算并进行调整. 该过程的伪代码算法 2 所示。

通过在一定环境中对任务的反复迭代学习实现对知识库的动态调整, 知识库中高频出现的连接关系将被加强, 低频出现的连接关系将被削弱, 子知识库被重新组织、更新和扩展, 同时一些与环境相关的新特征将会被总结生成新的元知识节点, 而较少被激活的元知识节点被边缘化 (处于被遗忘状态), 最终达到适应环境的目标. 智能架构中知识库的主要作用是建立环境特征和智能体执行动作之间的关联关系, 目前尚未有成熟的技术和方法能够完整实现上述的知识库构建、更新与扩展过程, 但在元知识生成、环境特征与元知识关联度计算、元知识间连接关系计算等方面, 现有方法中基于知识图谱^[31]、因果图^[32]等的知识表示方法等可以被借鉴。

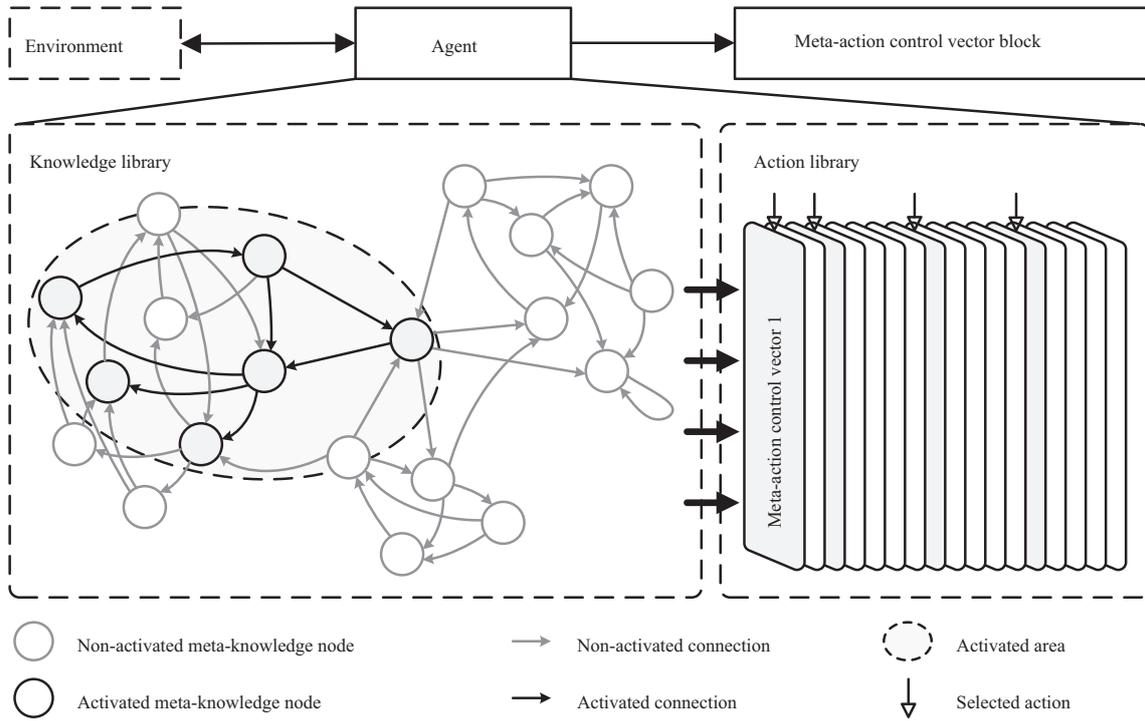


图 5 知识库与动作库的组织关系

Figure 5 Organizational relationship between the knowledge and action libraries

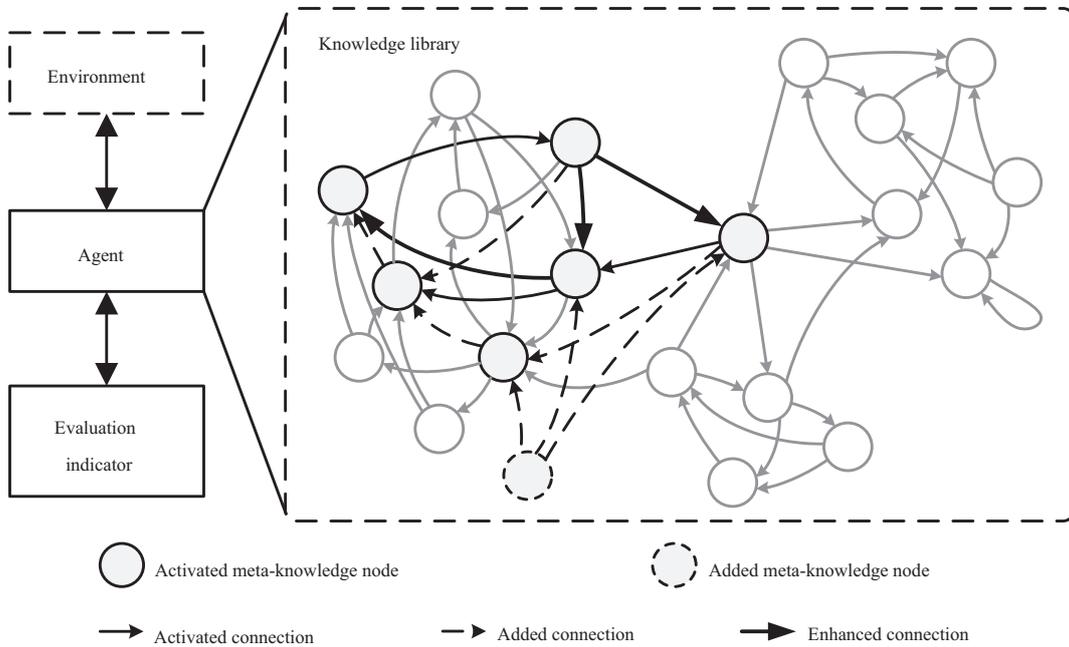


图 6 知识库更新、扩展过程

Figure 6 Update and expansion process of the knowledge library

表 1 无人车巡视采样任务描述

Table 1 Description of the surveying and sampling task

	Description
Task environment	150 × 150 units of discrete environment, where the rock weight is 1–9, the count is 300–500 (random), and the unmanned vehicle or each rock occupies 1 square unit
Task indicator	The surveying distance is 20–25 (absolute distance from the starting point, random); the sampling weight is 5–11 (random); the initial energy of unmanned vehicle is 50; the task ends when the task indicator is met or the energy is none
Unmanned vehicle	Four actions can be executed: forward, left 45°, right 45°, sampling; the vehicle can sample the rock directly ahead; the forward and steering actions cost 1 energy while the sampling action costs 8 energy; the vehicle's detection radius is 10 units; the vehicle can obtain the rocks' location and weight
Action library	Meta-action set {forward, left 45°, right 45°, sampling}; sub-action set {empty}
Knowledge library	Meta-knowledge set {rock, location, unmanned vehicle, distance, energy, sampling weight}; sub-knowledge set {empty}

算法 2 Update and expansion algorithm of the knowledge library

Initialization: Meta-knowledge set N_m , sub-knowledge set N_s , correlation α , threshold γ_N , $k_f = 0$;

Task execution:

- 1: Perceive the environment, extract features of the environment f , and calculate the correlation $\alpha_f : [\alpha_{f1}, \alpha_{f2}, \dots]$ between f and meta-knowledge nodes in $N_m : [n_{m1}, n_{m2}, \dots]$;
- 2: **for** α_{fi} in α_f **do**
- 3: **if** $\alpha_{fi} < \alpha$ **then**
- 4: **if** $k_{fi} > \gamma_N$ **then**
- 5: Generate new meta-knowledge n_{mf} corresponding to f and add it to N_m ;
- 6: $k_{fi} \leftarrow 0$;
- 7: **else**
- 8: $k_{fi} \leftarrow k_{fi} + 1$;
- 9: **end if**
- 10: **end if**
- 11: **end for**
- 12: Calculate the connection relationship between meta-knowledge in N_m and update N_s ;

4 智能架构实施示例

使用地外星球探测任务中无人车巡视采样任务说明具有学习和推理能力的无人系统智能架构的实现过程, 任务描述见表 1. 对应图 2 所示的智能体内信息流, 感知机根据无人车的探测结果提取岩石重量和坐标; 知识搜索模块组织元知识, 生成兴趣目标、任务完成度; 动作生成模块生成无人车要执行的元动作; 评估模块评价任务完成情况, 生成对应的奖励. 由于任务场景相对简单, 感知机、知识搜索模块和评估模块使用固定策略, 动作生成模块使用神经网络, 利用深度强化学习方法 (D3QN 算法^[33]) 训练智能体, 具体见表 2.

智能体训练的初始时刻, 由于没有对动作生成模块进行预训练, 智能体只能随机调用元动作执行任务, 经过一段时间训练后, 智能体首先学会前进动作以完成巡游任务, 接着智能体学会寻找岩石并抓取动作以完成采样任务. 训练过程的收敛曲线如图 7 所示, 经过约 1 万次的训练后, 任务成功率稳定在 90% 左右, 此时智能体能够较为高效率地完成该巡视采样任务. 训练后智能体的任务执行过程如

表 2 智能架构各模块功能描述
Table 2 Description of function modules

Module	Description
Perception	Obtain the rocks' location and weight within the detection range $[(x_1, y_1, m_1), (x_2, y_2, m_2), \dots]$
Knowledge search	Select the target rock (x, y, m) according to the current surveying distance S , sampling weight M , energy E ; if no target rock, select $(0, 0, 0)$; activate the executable actions (a_1, a_2, a_3, a_4) by setting the executable as 1 and non-executable as 0
Action generation	Use a three-layer full-connected network, where the number of nodes is $(10, 128, 4)$, the input is $(S, M, E, x, y, m, a_1, a_2, a_3, a_4)$, the output is the executed action, and the Softmax activation function is used in the output layer
Evaluation	Reward $1+E/10$ if the task is complete, while -1 otherwise in the output layer

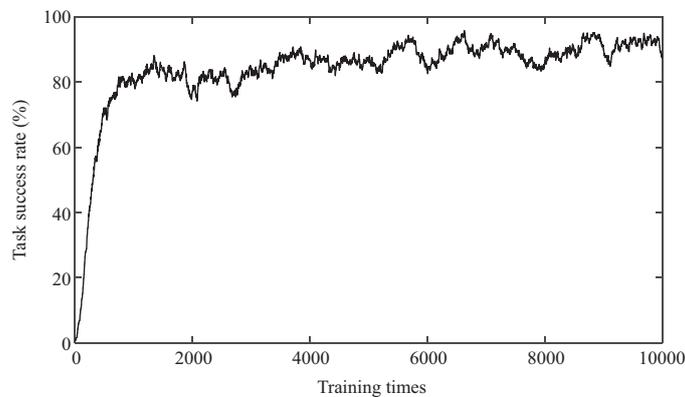


图 7 训练过程收敛曲线
Figure 7 Convergence curve of the training process

图 8 所示, 图中左上角标注了任务指标和实际完成情况, 其任务执行过程均包含了“发现目标、转向目标、抓取目标”的过程, 该寻找岩石并抓取的过程可作为 B 类子动作 (具体定义见 3.2 小节). 智能体还建立了 {环境特征}-{岩石、方位、距离、采集量}-{抓取} 之间的联系, 其中 {岩石、方位、距离、采集量}-{抓取} 可作为新的子知识, 如表 3 所示. 分析无人车的移动路径可知, 在巡游采样任务驱动下, 智能体先向前运动执行巡游子任务, 巡游过程中探测到符合要求的岩石后靠近并抓取完成采样子任务, 再完成剩余的巡游子任务, 智能体在未知环境下通过对元动作的有效组织和利用, 自主生成了任务执行策略. 可以看出, 通过有效的环境探索、推理与学习, 无人车自主产生出 B 类子动作和新的子知识, 并自动生成了任务执行策略, 验证了所提出的智能架构可以实现知识的生成和任务的自主执行.

要指出的是, 本示例旨在说明智能架构的运作机制, 通过一个简单的设计和一个成熟的强化学习方法, 验证了在该架构下智能体可学习、知识可生产、策略可推理, 在学习方法、知识表达等方面还需继续深入研究.

5 结论

为提高航天器对未知空环境和复杂空间任务的适应能力, 本文提出具有学习和推理能力的无人系统智能架构, 定义智能架构由智能体、环境和任务 3 部分组成, 使用动作库和知识库描述智能体, 并

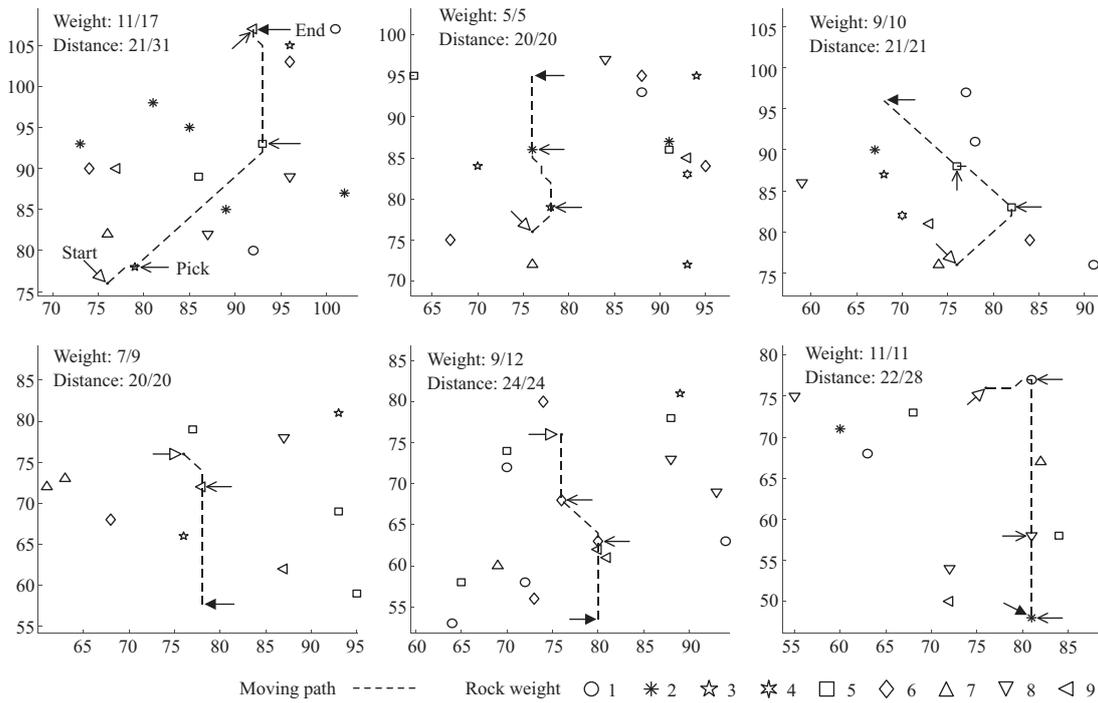


图 8 训练后的任务执行过程

Figure 8 Task execution process after training

表 3 生成的子动作和子知识

Table 3 Generated sub-action and sub-knowledge

	Description
Added sub-action	Search rock and sample
Added sub-knowledge	{rock, location, distance, sampling weight}-{sampling}

能够在与环境 and 任务交互过程中在线更新; 针对智能架构中动作库和知识库的构建、更新与扩展方法, 智能体的学习和推理机制等关键问题, 给出解决方案和技术途径; 最后使用无人车巡视采样任务说明了智能架构的实现过程, 结果表明, 通过在不同环境下训练的无人车自主产生出 B 类子动作和新的子知识, 并自动生成了任务执行策略, 验证了所提出的智能架构具有学习和推理能力, 可以实现知识的生成和任务的自主执行. 后续, 在该智能架构的框架下, 针对各功能模块的构建以及智能架构的推广应用等方面的研究还需继续深入开展.

参考文献

- 1 Giaimo F, Berger C. Design criteria to architect continuous experimentation for self-driving vehicles. In: Proceedings of 2017 IEEE International Conference on Software Architecture (ICSA 2017), Gothenburg, 2017. 203-210
- 2 Pope A P, Ide J S, Micovic D, et al. Hierarchical reinforcement learning for air-to-air combat. In: Proceedings of the 2021 International Conference on Unmanned Aircraft System (ICUAS 21), Athens, 2021
- 3 杨嘉墀. 中国空间计划中智能自主控制技术的发展. 见: 1995 年中国控制会议论文集 (上), 1995. 1-5
- 4 中华人民共和国国务院新闻办公室. 《2016 中国的航天》白皮书. 中国航天, 2017, 1: 10-17
- 5 国务院. 国务院印发新一代人工智能发展规划的通知. 国务院公报, 2017, 22: 7-21

- 6 Guo H. Research on the development strategy of intelligent unmanned system. *Unmanned Syst Technol*, 2020, 3: 1–11 [郭行. 智能无人系统发展战略研究. *无人系统技术*, 2020, 3: 1–11]
- 7 Dafarra S, Bertrand S, Griffin R J, et al. Non-linear trajectory optimization for large step-ups: application to the humanoid robot Atlas. In: *Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020
- 8 Luo J, Yan B, Wood K. InnoGPS for data-driven exploration of design opportunities and directions: the case of google driverless car project. *J Mech Des*, 2017, 139: 111416
- 9 Yan D, Zhou N E. The applications and prospects of CH UAV systems. *Comput Eng Software*, 2018, 39: 117–122 [闫东, 周乃恩. 彩虹无人机系列应用及展望. *软件*, 2018, 39: 117–122]
- 10 Tang Y G, Wang J, Lu Y, et al. Parametric design method and experimental research on Haidou full-depth ocean autonomous and remotely-operated vehicle. *Robot*, 2019, 41: 697–705 [唐元贵, 王健, 陆洋, 等. “海斗号” 全海深自主遥控水下机器人参数化设计方法与试验研究. *机器人*, 2019, 41: 697–705]
- 11 冯健翔. 人工智能及其航天应用概论. 北京: 宇航出版社, 1999. 21–22
- 12 Rayman M D, Varghese P, Lehman D H, et al. Results from the Deep Space 1 technology validation mission. *Acta Astronaut*, 2000, 47: 475–487
- 13 Di K C, Ge Z J. A brief review of Spirit’s six years of Mars roving and scientific discoveries. *J Remote Sensing*, 2011, 15: 651–658 [邸凯昌, 葛之江. 勇气号火星车六年探测征程及科学发现简述. *遥感学报*, 2011, 15: 651–658]
- 14 Farley K A, Williford K H, Stack K M, et al. Mars 2020 mission overview. *Space Sci Rev*, 2020, 216: 142
- 15 Wu H X, Hu J, Xie Y C. Spacecraft intelligent autonomous control: past, present and future. *Aerospace Control Appl*, 2016, 42: 1–6 [吴宏鑫, 胡军, 解永春. 航天器智能自主控制研究的回顾与展望. *空间控制技术与应用*, 2016, 42: 1–6]
- 16 Yuan L, Wang S Y. A review on development of intelligent health management technology for spacecraft control systems. *Acta Aeronaut ET Astronaut Sin*, 2021, 42: 525044 [袁利, 王淑一. 航天器控制系统智能健康管理技术发展综述. *航空学报*, 2021, 42: 525044]
- 17 钟兴, 巩加龙, 王春月, 等. 吉林一号广域高频次高分辨智能遥感应用. *卫星应用*, 2021, 10: 31–39
- 18 Wu W R, Liu J Z, Tang Y H, et al. China Lunar Exploration Program. *J Deep Space Explor*, 2019, 6: 405–416 [吴伟仁, 刘继忠, 唐玉华, 等. 中国探月工程. *深空探测学报*, 2019, 6: 405–416]
- 19 Zhao Y, Wang X L, Huang X Y, et al. Tianwen-1 lander guidance navigation and control system for mars soft landing. *Aerospace Control Appl*, 2021, 47: 48–57 [赵宇, 王晓磊, 黄翔宇, 等. 天问一号火星软着陆制导、导航与控制系统. *空间控制技术与应用*, 2021, 47: 48–57]
- 20 Yuan L, Huang H. Current trends of spacecraft intelligent autonomous control. *Aerospace Control Appl*, 2019, 45: 7–18 [袁利, 黄煌. 空间飞行器智能自主控制技术现状与发展思考. *空间控制技术与应用*, 2019, 45: 7–18]
- 21 Chiang C H, Chen L H. A self-adaptive intelligent control system with hierarchical architecture. *J Inform Sci Eng*, 2015, 31: 1027–1049
- 22 Murphy R R. *Introduction to AI Robotics*. Cambridge: The MIT Press, 2000. 274–277
- 23 Hassani K, Lee W S. A software-in-the-loop simulation of an intelligent microsatellite within a virtual environment. In: *Proceedings of 2013 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2013. 31–36
- 24 Narayan P, Wu P, Campbell D A, et al. An intelligent control architecture for unmanned aerial systems (UAS) in the National Airspace System (NAS). In: *Proceedings of the 2nd Australasian Unmanned Air Vehicle Systems Conference*, 2007
- 25 Qureshi F, Terzopoulos D. Intelligent perception and control for space robotics. *Machine Vision Appl*, 2008, 19: 141–161
- 26 Shi W, Wang K W, Yu C P, et al. Artificial intelligence security in multiple-unmanned system cooperation. *Strategic Study CAE*, 2021, 23: 82–89 [施文, 王楷文, 俞成浦, 等. 多无人系统协同中的人工智能安全探索. *中国工程科学*, 2021, 23: 82–89]
- 27 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge: The MIT Press, 2018. 45–66
- 28 Wang R P, Gao Z H. Research on decision system in air combat simulation using maneuver library. *Flight Dynamics*, 2009, 27: 72–79
- 29 Yang Y M, Nie B, Fang H G, et al. Planning collision-free grasping actions of virtual human. *J Comput-Aided Design Comput Graph*, 2014, 8: 1362–1373

- 30 Won J, Gopinath D, Hodgins J. Control strategies for physically simulated characters performing two-player competitive sports. *ACM Trans Graph*, 2021, 40: 1–11
- 31 Wang Q, Mao Z, Wang B, et al. Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans Knowl Data Eng*, 2017, 29: 2724–2743
- 32 Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books Press, 2018. 1–30
- 33 Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning. 2016. ArXiv:1511.06581

Intelligent architecture of space unmanned system with learning and reasoning capabilities

Huang HUANG^{1,2}, Mou LI^{1,2*}, Lei LIU^{1,2}, Liang TANG^{1,2}, Hao LIU^{1,2}, Xinru XIE^{1,2}, Nailong LIU^{1,2}, Chunling WEI^{1,2}, Yan XING^{1,2}, Tiantian JIANG^{1,2}, Haidong HU^{1,2}, Yafei CHANG^{1,2}, Yong HU^{1,2} & Mengfei YANG³

1. *Beijing Institute of Control Engineering, Beijing 100190, China;*

2. *Key Laboratory of Space Intelligent Control Technology, Beijing 100094, China;*

3. *Chinese Academy of Space Technology, Beijing 100094, China*

* Corresponding author. E-mail: limou2333@163.com

Abstract Spacecraft represented by satellites, manned spaceships, and deep space probes are typical unmanned systems that are developing from automation to autonomy. To improve the adaptability and intelligent autonomy level of spacecraft to unknown space environments and complex space tasks, an intelligent architecture of space unmanned system with learning and reasoning capabilities is designed and the architectural components and functions are proposed. Then, the operating mechanism of learning and reasoning capabilities in the architecture are analyzed, and the solutions for key technologies, such as the construction, update, and expansion methods of the action and knowledge libraries, are proposed. Finally, through an example of the extraterrestrial surveying and sampling task, the process of new knowledge generation, knowledge-based reasoning, and autonomous task execution under this intelligent architecture, as well as the interaction between the knowledge library, environment, task, and action library, are explained.

Keywords learning and reasoning capabilities, space unmanned system, intelligent architecture, action library, knowledge library