自主智能体动态能力自我评估

Nicholas Conlon
nicholas.conlon@colorado.edu 科罗拉多大学博尔德分校 博尔德，科罗拉多州，美国

Nisar R. Ahmed
nisar.ahmed@colorado.edu 科罗拉多大学博尔德分校 博尔德，科罗拉多州，美国

Daniel Szafir
daniel.szafir@cs.unc.edu 北卡罗来纳大学教堂山分校 教堂山，北卡罗来纳州，美国

arXiv:2303.01646v1 [cs.RO] 3 Mar 2023

摘要

随着自主机器人被部署到日益复杂的环境中，平台退化、环境不确定性以及偏离经验证的运行条件，都会使人类伙伴难以了解机器人能力与局限。在动态且不确定的环境中对自身能力进行自我评估，是机器人成功实现人机协作的关键下一步。本文介绍并评估一种面向自主智能体的事件触发广义结果评估（ET-GOA）算法，该算法可在执行过程中动态评估任务置信度。该算法利用一种快速在线统计检验，对智能体的观测结果及其模型预测进行分析，以决定何时需要进行能力评估。我们提供了在模拟配送任务中使用 ET-GOA 生成能力报告的实验结果，并提出了自我评估智能体未来的研究方向。

CCS 概念

• **计算机系统组织** → 机器人自主性；• **计算方法** → 机器学习；《人工智能》；
• **以人为中心的计算** →交互技术。

关键词

人机协作，机器人自我评估

ACM 参考文献格式：
Nicholas Conlon、NisarR. Ahmed 和 Daniel Szafir。2023年。《自主智能体的动态能力自我评估》。收录于*ACM/IEEE HRI 2023*人机团队协作中的可变自主性研讨会论文集（*HRI 2023*人机团队协作中的可变自主性研讨会）。ACM，纽约州纽约，美国，5页。https://doi.org/X

1 引言

自主自我评估是推动复杂机器人部署的关键。设想这样一个场景：灾害发生后，一支搜索与救援（SAR）队伍正在运送急需的物资。环境十分危险，因此队伍决定使用半自主机器人在环境中导航并运送物资。队伍对机器人的依赖取决于他们对其能力的判断。然而，如果队伍对机器人能力的判断与机器人实际的能力与局限之间存在不匹配，队伍可能会无意中让机器人超越其能力边界 [13]。为了做出恰当的任务分配决策（例如选择安全的运送地点、进行控制权交接、选择自主级别），SAR 队伍必须了解机器人的能力，以及这种能力在任务期间可能发生的变化。

在近期的研究中，我们发现，报告先验 自我评估任务成功置信度的智能体，有助于使操作员感知与实际机器人能力相一致，从而提升可变自主导航任务中的决策质量和表现[8]。然而，在上文所述的搜索与救援场景这类动态且不确定的环境中，由于坠落碎片或障碍物等环境变化，先验 置信度评估可能很快就会过时。在本研究中，我们提出了一种名为事件触发广义结果评估 (ET-GOA)的算法，使自主智能体能够持续监测能力的现场 变化，并在必要时针对显著改变结果可能性的事件进行自我评估并报告。我们证明，该方法能够捕捉对任务成功产生积极或消极影响的动态环境变化；在自主决策中使用 ET-GOA 置信度评估，可显著提升任务性能。

2 背景与相关工作

能力自我评估使自主智能体能够根据任务约束和环境条件评估自身能力与局限性。这些关键信息可用于改进内部决策，和/或传达给人类伙伴以改进外部决策。任务或任务执行前（先验）自我评估使自主智能体能够在执行任务或任务之前评估自身能力。这些方法通常依据仿真 [2, 15]或以往经验 [10]计算智能体的自信度。我们近期的研究表明，报告先验 自我评估有助于更好地选择依赖程度 [7]，并提升表现和信任 [8]。然而，在动态环境中，由于执行前无法考虑的因素（例如环境变化、任务变化或与其他智能体的交互），先验 评估难以准确预测智能体的置信度。在线（定期）运行先验 方法或许能够捕捉能力的动态变化。然而，如果能力实际上并未发生变化，此类评估可能会浪费计算资源；对于某些类型的决策智能体而言，评估成本也可能过高 [1, 6, 11]。

任务期间（原位）自我评估使自主智能体能够在任务执行期间评估（或重新评估）其能力。在线行为分类等常用方法可以识别不良行为，并在发生故障时触发智能体停止运行并寻求帮助 [9, 17, 19]。这些方法虽然能够捕捉

特此免费授予许可，允许为个人或课堂使用而制作本作品的部分或全部内容的数字或纸质副本，但不得为营利或商业利益而制作或分发副本，且副本首页必须附有本声明和完整引文。必须尊重本作品中第三方组件的版权。其他用途请联系所有者/作者。*HRI 2023*人机团队协作中的可变自主性研讨会，*2023*年*3*月*13*日，瑞典斯德哥尔摩© 2023年版权归所有者/作者所有。ACM ISBN X。
https://doi.org/X

能力的动态变化，但需要良好（有能力）和不良（无能力）行为的示例，而在许多现实应用中，这些示例可能难以获取，甚至根本无法获取。另一种原位自我评估方法是监测智能体当前状态的特征。例如，Gautam 等人开发了一种监测偏离设计假设情况的方法 [11],，而 Ramesh 等人则使用机器人的“生命体征”来监测其在任务执行期间的健康状况 [16]。这两种方法都能提供智能体在特定时刻状态的宝贵即时快照，可用于在线指示表现是否正在退化；然而，两者都无法预测更高层次的任务能力（例如，这种退化是否真的会影响任务结果？）。相比之下，我们提出一种原位自我评估方法，通过监测智能体的预测与观测之间的一致性来实现混合方案；当两者出现偏差时，则使用一种准确的先验方法触发任务置信度的（重新）评估。

3 任务置信度的动态自我评估

3.1 何时评估能力

为避免不必要的评估并节省机载计算资源，我们采用事件触发的方法进行原位自我评估。只有在有证据表明智能体的任务置信度发生变化时，我们的算法才会评估置信度。

检测此类变化的一种有前景的方法是惊奇指数（SI）。给定概率模型 [20]，SI 定义为比观测事件更极端（或概率更低）的事件的概率总和。对于给定事件 $e \in E$，SI 的计算方法是对分布 $p(E)$ 中更极端事件的概率求和：

$$SI(e, p(E)) = \int_{p(E)<p(e)} p(E)dE \quad (1)$$

SI 可以理解为，在给定一组可能事件 E 的情况下，观测值 e 有多（不）相容。这与更广为人知的基于熵的惊讶度 [3, 4]类似；不过，基于熵的惊讶度没有上界，而惊讶指数的取值介于零（最令人惊讶）和一（最不令人惊讶）之间。SI 也与尾概率或 p值类似：在假设e 来自分布 p（E）的前提下，较大的 p值（较大的 SI）表明有力证据支持 e 很可能来自 p（E）；而较小的 p值则表明有力证据支持相反的结论。

在这项工作中，我们关注的是确定智能体何时应重新评估其任务置信度。我们提出，计算智能体观测到的状态 s_t 相对于该智能体模型预测 $p(\hat{s}_t)$的 SI，并在 SI 低于阈值 δ时触发重新评估。本质上，我们会监测智能体模型针对给定任务的质量，并在质量下降时触发评估。由于 s_t 的某些方面可能比其他方面与能力更相关，我们会针对状态边缘分布的一个子集计算 SI。

3.2 如何评估能力

为评估能力，我们采用广义结果评估（GOA） [6]。给定概率世界模型 M，GOA 通过展开状态预测 $p_M(s_{t+1}|s_t,\ a_t)$来模拟任务执行。请注意，a M可以采用基于蒙特卡洛的规划器[14]的形式，黑盒神经网络 [6, 12],或类似模型。随后，GOA 会分析状态预测，并计算智能体达到优于某个目标结果阈值 Z 的结果的置信度余量。目标结果的示例包括击中的陨石坑数量（我们希望越少越好）或投递的包裹数量（我们希望越多越好）。置信值可以报告为原始数值 $\in (0,1)$，也可以映射为表示置信度的语义标签，例如极有可能、有可能、不太可能、极不可能。对于后文所述的实验，我们使用置信度的原始数值。

3.3 事件触发广义结果评估算法

我们将基于意外的动态自我评估方法称为事件触发广义结果评估（ET-GOA）。该算法见算法 1，可分为两个部分：（1）任务执行之前；（2）任务执行期间。任务执行之前（第 *1–5* 行）：第 1 行的输入包括模型**M**、任务规范**T**、一组结果阈值**Z**（每个结果对应一个），以及一组意外阈值 δ（每个感兴趣的状态边缘分布对应一个）。接下来（第 2 行），使用模型**M**，根据初始状态 s_0模拟任务**T**的执行。这会得到一组预测状态分布 $[p(\hat{s}_t)]_{t=0:N}$，每个时间步 t对应一个。我们进一步将给定时间步的状态分布拆分为 K 边缘分量。例如，如果我们想在 SI 触发器中使用 $x, y,$ 和z位置，则 $K = 3$和 $p(\hat{s}_t)$ 将拆分为一组边缘概率分布 $[p(\hat{s}_{t,x})\ p(\hat{s}_{t,y})\ p(\hat{s}_{t,z})]$。这一额外的边缘化步骤在算法中是隐含的，但值得注意。随后，将每个时间步的预测边缘分布存储在经验缓冲区中（第 3 行），再用于计算初始广义结果评估（第 4 行），并可将评估结果报告给操作员（第 5 行）。

算法 1： 事件触发广义结果评估-算法

```
1  算法 ET-GOA(M, T, Z, δ)：
2      [p(ŝ_1) ..., p(ŝ_N)] ← 模拟 M(T s_0)
3      exp_缓冲区 ← [p(ŝ_1) ..., p(ŝ_N)]
4      goa ← GOA(exp_缓冲区,Z)
5      GOA 报告
6      对于 ( t 中 1 : N )
7          s_t ←接收_状态_观测(t)
8          p(ŝ_t) ← exp_缓冲区(t)
9          si_min = 分钟_{i=1:K} SI(s_{t,i}, p(ŝ_{t,i}))
10         如果 si_min ≤ δ 那么
11             [p(ŝ_{t+1}) ..., p(ŝ_N)] ← 模拟 M(T s_t)
12             exp_缓冲区 ← [p(ŝ_{t+1}) ..., p(ŝ_N)]
13             goa ← GOA(exp_缓冲区，Z)
14             报告 GOA
15         else
16             继续
```

在任务执行期间（第 *6–16* 行）：智能体在时间 t 观测状态 s_t（第 7 行）。随后，它从经验缓冲区中获取时间的状态分布（即预测值）（第 8 行）。

接下来，算法计算 si_{min}，即在给定预测边缘分布 $p(\hat{s}_{t,i})$的情况下，各个 K 观测状态边缘分布$s_{t,i}$的 SI 最小值（第 9 行）。如果 si_{min} 低于 δ，则表示收到异常或令人意外的状态观测，需要重新评估置信度（第 10 行）。此时，会根据 M 模拟一组新的预测状态分布（第 11 行），并将其保存到经验缓冲区中（第 12 行）。随后，利用更新后的经验缓冲区计算新的自我评估（第 13 行），并报告给操作员（第 14 行）。 si_{min} 高于 δ则表明智能体的预测与其观测一致，此时无需更新置信度（第 16 行）。这个循环（第 6 行）会在整个任务期间持续进行，将预测的状态边缘分布与实际观测进行比较，并在必要时重新计算智能体的任务置信度并报告更新。

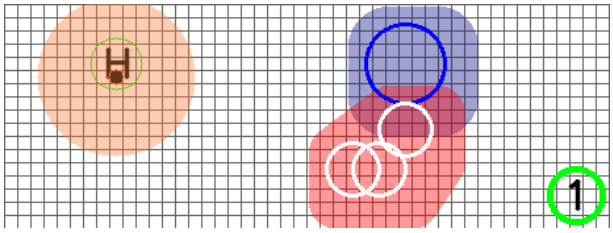

图 1：示例环境，展示智能体的位置和视场（橙色）、目标区域（绿色）、尘埃区（蓝色圆圈）及陨石坑（白色圆圈）的真实位置。障碍物以蓝色和红色突出显示，以增强视觉对比度。

4 实验

我们通过两项仿真实验评估了 ET-GOA。第一项实验考察 ET-GOA 对任务性能的影响；第二项实验考察 ET-GOA 捕捉任务难度变化的能力。

4.1 配送场景概述

我们的实验场景基于第 1 节中用于说明问题的搜索与救援示例：一个智能体负责将货物安全送达三个目标之一。环境中有两类难以避开的障碍物：陨石坑和尘埃区。驶过陨石坑会损坏智能体；如果导航过程中撞上足够多的陨石坑，智能体就会被视为损坏，并导致配送任务失败。尘埃区会降低传感器性能，并给智能体的状态转移动态引入噪声。尘埃区通常位于陨石坑附近，因此智能体一旦进入尘埃区，撞上陨石坑的可能性就会增加。为了模拟真实部署中可能发生的环境变化，任务执行期间，新的障碍物可能会在随机位置（智能体所在位置除外）出现。

该环境是一个自定义的 OpenAI Gym 环境 [5]。智能体被建模为一个离散状态/动作马尔可夫决策过程，其状态空间为 $s = (s_x, s_y, s_c, s_z)$，由智能体的位置（x，y）以及其传感器视场（FOV）内陨石坑（s_c）和尘埃区（s_z）的数量构成。传感器视场被建模为半径为 10 个网格方格的全向视场。二维环境的总大小为 50x50 个网格方格。我们使用 Q学习 [18]，针对每个目标分别训练了一个策略。训练期间不设置任何障碍物，以防止智能体学习如何克服环境中的困难。

用于自我评估的世界模型 M 是环境的副本，具有完全相同的状态转移函数，但只包含已知的陨石坑和尘埃区。智能体选择评估置信度最高的目标。若置信度相同，智能体则选择距离最近的目标。图1展示了一个环境示例。

我们评估了两种不同的环境：静态和动态。在静态环境中，智能体预先已知陨石坑和尘埃区的位置，并且这些位置在整个任务执行过程中保持不变。在动态环境中，陨石坑和尘埃区的位置起初是已知的，但在某个预定时间发生变化，以模拟先前生成的机载导航地图突然变得过时。

4.2 假设

我们提出了三个假设：（1）在静态环境中，ET-GOA 和 GOA 的表现同样出色，且都优于随机目标选择；（2）在动态环境中，ET-GOA 的表现优于 GOA 和随机选择；（3）ET-GOA 能够捕捉任务难度的正向和负向变化。针对假设 1 和 2，我们分析了智能体表现（投送次数）；针对假设 3，我们分析了相对于任务难度变化所报告的置信度。

4.3 表现提升

我们的第一个实验用于验证前两个假设。在 $t = 0$，我们向智能体提供了所有障碍物的位置。在动态条件下，我们于 $t = 10$ 更改障碍物的位置，但不告知智能体。我们考虑了三种条件：不进行评估、*GOA*和*ET-GOA*。不进行评估条件下不使用任何能力评估。相反，在 $t = 0$，智能体随机选择目标并直接向其导航。*GOA* 条件采用了 [6]中讨论的标准广义结果评估分析。在 $t = 0$，智能体根据公式 2 选择 GOA 置信度最高的目标并向其导航 $g \in G$。

$$g = \arg\max_{i \in G} GOA_i \tag{2}$$

ET-GOA 条件使用了第 3.3 节所讨论的 ET-GOA 算法。我们使用两个状态边缘分布作为触发条件：s_c（以及 $\hat{s}_c$），即智能体视场中实际（以及预测的）可见陨石坑；以及 s_z（以及 $\hat{s}_z$），即智能体视场中实际（以及预测的）可见尘埃区。这实质上计算了预期障碍物位置（根据初始位置信息）与智能体穿越环境时观测到的“地面上”障碍物位置之间的意外程度。我们选择这些特定的状态边缘分布，是因为它们与现代机器人中的传感器能力相一致。此外，查看智能体传感器视场中的意外观测，使算法能够在智能体实际接触到可能危险的障碍物之前触发重新评估。如果任一状态边缘分布的最小 SI 小于以下值，算法便会触发重新评估：

状态边缘分布小于 $\delta = 0.05$:

$$\min(SI(s_c, \hat{s}_c), SI(s_z, \hat{s}_z)) < 0.05$$

随后，智能体根据公式 2 选择目标。智能体直接导航至该目标，直到到达目标或触发重新评估并选择新目标。对于每种条件，智能体都尝试了 100 次配送任务。

4.3.1 结果。我们发现，环境对投送次数有显著的主效应（$t(598) = 4.65$，$p < 0.0001$），表明静态环境比动态环境更容易应对，这符合预期。在静态环境中，我们发现每种报告条件对投送次数都有显著影响（$F(2, 297) = 43.5$，$p < 0.0001$）。使用 Tukey HSD检验进行事后分析发现，与随机条件相比，ET-GOA 条件下的投送次数显著增加（$p = 0.0001$）；与随机条件相比，GOA 条件下的投送次数也显著增加（$p < 0.0001$）。静态环境中 GOA 与 ET-GOA 之间没有差异，这符合预期。在动态环境中，我们发现每种报告条件对投送次数都有显著影响（$F(2, 297) = 44.1$，$p < 0.0001$）。使用 Tukey HSD检验进行事后分析发现，与随机条件（$p = 0.0001$）和 GOA 条件（$p < 0.0001$）相比，ET-GOA 条件下的投送次数均显著增加。这些结果证实了我们的第一和第二个假设，详见图2。

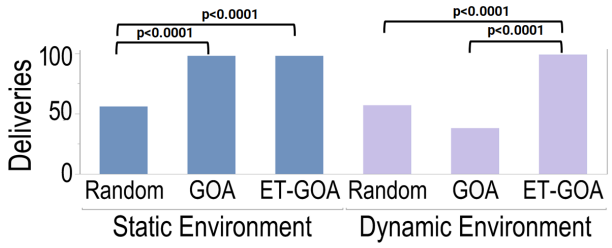

图 2：每种条件下进行 100 次投送尝试的结果图显示，在静态环境中，ET-GOA 的表现显著优于随机方法（左）；在动态环境中，其表现显著优于随机方法和 GOA（右）。

4.4 检测难度变化

我们的第二个实验用于验证第三个假设。在此，我们评估了 ET-GOA 对影响任务难度的环境变化的捕捉效果。为进行评估，我们让智能体在两种条件下前往一个固定的目标位置。第一种条件称为 $easy \rightarrow hard \rightarrow easy$，初始时没有障碍物；在时间步 10 随机添加障碍物，然后在时间步 30 删除所有障碍物。第二种条件称为$hard \rightarrow easy \rightarrow hard$，初始时有随机分布的障碍物；在时间步 10 删除所有障碍物，然后在时间步 30 添加新的障碍物。添加障碍物会增加智能体的任务难度，反之，删除障碍物则会降低任务难度。时间步零时的障碍物为智能体已知，而在时间步 10 和 30 添加或删除的障碍物则必须在现场观察到。我们在每种条件下运行了 100 个回合，并记录初始评估以及每次添加或删除事件后 ET-GOA 的评估结果。

4.4.1 结果。我们观察到，在 $easy \rightarrow hard \rightarrow easy$ 和 $hard \rightarrow easy \rightarrow hard$ 任务与初始评估时、第一次环境变化后（$t(99) = 27.7$，$p < 0.0001$）以及第二次环境变化后（$t(99) = 37.5$，$p < 0.0001$），智能体置信度存在显著差异。这证实了我们的第三个假设：ET-GOA 能够捕捉任务难度受到的正面和负面影响。结果图见图3。

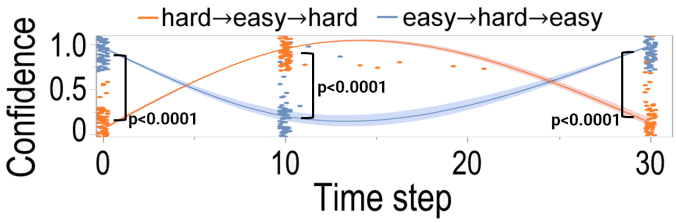

图 3：结果图显示，ET-GOA 捕捉到了任务难度的变化。$hard \rightarrow easy \rightarrow hard$ 任务以橙色表示，$easy \rightarrow hard \rightarrow easy$ 任务以蓝色表示。任务难度变化事件发生在 $t = 10$ 和 $t = 30$。实线表示任务置信度的均值和标准差。

5 结论与未来工作

本文提出了一种用于事件触发的广义结果评估的算法，可在动态且不确定的环境中计算自主智能体的实时任务置信度。ET-GOA 根据智能体预测状态和实际状态之间的惊讶指数，决定何时评估任务置信度。我们在静态和动态环境中评估了 ET-GOA 执行投送任务的效果，发现与基线方法相比，该方法显著提升了性能。我们还发现，ET-GOA 能够捕捉智能体置信度的变化，从而反映任务难度的变化。也就是说，我们的方法可以判断任务何时变得更难或更容易。下一步，我们将通过实地平台和人类受试者研究验证 ET-GOA。我们推测，ET-GOA 的应用将帮助操作员在决定是否依赖自主机器人时做出更好的决策，从而提升表现并降低工作负荷。

ET-GOA 能够使自主机器人提供有关其对任务成功的“实地”置信度、置信度变化情况及其原因等关键信息。我们相信，这对人机团队将极具价值，尤其是对那些在高风险且不确定的环境中工作的团队而言；在这些环境中，人类操作员需要就任务执行、自主程度和/或控制做出关键决策。

致谢

本研究得到美国国防高级研究计划局（DARPA）合同编号 HR001120C0032 的支持。本文所表达的任何观点、研究结果、结论或建议均属于作者个人意见，不一定代表 DARPA 的立场。

参考文献

[1] Aastha Acharya、Rebecca Russell 和 Nisar R. Ahmed。2022年。使用深度生成模型对自主智能体进行能力评估。载于*2022*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。8211–8218。https://doi.org/ 10 1109 47612 2022 9981991 。/IROS。。

[2] Paola Ardón、Eric Pairet、Yvan Petillot、Ronald Petrick、Subramanian Ramamoor- thy 和 Katrin Lohan。2020年。抓取可供性迁移的自我评估。载于*2020*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE。https://doi.org/ 10 1109 45743 2020 9340841 。/IROS。。

[3] Pierre Baldi 和 Laurent Itti。。比特与惊奇：惊奇的贝叶斯理论及其在注意力中的应用。《神经网络》23, 5（2010），649–666。https://doi.org/ 10 1016 2009 12 007 。/ 2010 j.neunet。。

[4] William A. Benish。。相对熵作为诊断信息的度量。《医学决策》19, 2（1999），202–206。https://doi.org/10.1177/ 0272989X9901900211 1999

[5] Greg Brockman、Vicki Cheung、Ludwig Pettersson、Jonas Schneider、John Schul- man、Jie Tang 和 Wojciech Zaremba。2016年。OpenAI Gym。*CoRR* abs/1606.01540（ 2016 1606 01540 1606 01540 ）。arXiv：。

[6] http://arxiv.org/abs/。Nicholas Conlon、Aastha Acharya、Jamison McGinley、Trevor Slack、C. Alexander Hirst、Mitchell Hebert、Chris Reale、Eric Frew、Rebecca Russell 和 Nisar Ahmed。2022年。使用深度强化学习推广自动驾驶车辆的能力自我评估。载于*AIAA SciTech*论坛。AIAA。[7] Nicholas Conlon、Daniel Szafir 和 Nisar Ahmed。2022年。探究机器人熟练度自我评估对信任和表现的影响。载于*AAAI*春季研讨会系列论文集：闭合评估闭环：在人机协作中传达熟练度与意图。AAAI。[8] Nicholas Conlon、Daniel Szafir 和 Nisar Ahmed。2022年。“我确信这会以糟糕的结果告终”：人机协作中的机器人熟练度自我评估。载于*2022*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。 2127 2134 10 1109 47612 2022 9981653 –。https://doi.org/。/IROS。。

[9] Maria Fox、Malik Ghallab、Guillaume Infantes 和 Derek Long。2006年。通过学习的隐马尔可夫模型进行机器人内省。《人工智能》170, 2 2006 59 113 10 1016 2005 05（）$_{007}$ –。https://doi.org/。/j.artint。。[10]

Tyler Frasca、Evan Krause、Ravenna Thielstrom 和 Matthias Scheutz。2020年。“你能做到这个吗？”自主机器人在任务执行前、执行中和执行后的自我评估对话，载于《2020年评估、解释和传达机器人在人机协作中熟练度研讨会》论文集。*arXiv*电子预印本。

[11] Alvika Gautam、Tim Whiting、Xuan Cao、Michael A. Goodrich 和 Jacob W. Crandall。2022年。一种设计知晓自身局限的自主机器人的方法。载于*2022*年国际机器人与自动化会议（*ICRA*）。IEEE，美国宾夕法尼亚州费城， 121 127 10 1109 46639 2022 9812030 – 。

[12] https://doi.org/ 。 /ICRA 。 。David Ha 和 Jürgen Schmidhuber。 。循环世界模型促进策略演化。载于神经信息处理系统进展，S. Bengio、H. Wallach、H. Larochelle、K. Grauman、N. Cesa-Bianchi 和 R. Garnett（编），第31卷。Curran Associates, Inc. https://proceedings.neurips.cc/paper/2018/file/ 2018 2de5d16682c3c35007e4e92982f1a2ba-Paper.pdf [13] Andrew R Hutchins、Mary L Cummings、Mark Draper 和 Thomas Hughes。2015年。通过能力边界呈现自主系统的自信度。载于人因与工效学学会年会论文集，第59卷。SAGE Publications Sage CA：美国加利福尼亚州洛杉矶，Sage，279–283。[14] Brett Israelsen。2019年。自主系统中的算法保证与能力边界自我评估。博士学位论文。科罗拉多大学博尔德分校，博尔德。[15] Brett Israelsen、Nisar Ahmed、Eric Frew、Dale Lawrence 和 Brian Argrow。2019年。通过对决策过程进行元分析提升自主系统的机器自信度。载于应用人因与工效学国际会议。Springer，Springer，213–223。[16] Aniketh Ramesh、Rustam Stolkin 和 Manolis Chiou。2022年。机器人生命体征与机器人健康：系统性量化不利条件下机器人运行时表现退化。*IEEE*机器人与自动化快报7，4（2022年），10729–10736。https://doi.org/10.1109/LRA.2022.3192612[17] Juan Rojas、Shuangqi Luo、Dingqiao Zhu、Yunlong Du、Hongbin Lin、Zhengjie Huang、Wenwei Kuang 和 Kensuke Harada。2017年。通过基于力旋量的动作语法实现在线机器人内省。载于*2017*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，5429–5436。ISSN：2153-0866。[18] Christopher JCH Watkins 和 Peter Dayan。1992年。Q学习。机器学习 8，3（1992年），279–292。[19] Hongmin Wu、Hongbin Lin、Yisheng Guan、Kensuke Harada 和 Juan Rojas。2017年。利用贝叶斯非参数向量自回归隐马尔可夫模型进行机器人内省。载于*2017*年*IEEE-RAS*第*17*届国际人形机器人会议（*Humanoids*）（英国伯明翰）。IEEE，882–888。https: // doi.org/ 10 1109 2017 8246976 . /HUMANOIDS. 。Adam Zagorecki、Marcin Kozniewski 和 Marek J. Druzdzel。2015年。作为置信度度量的惊讶指数近似。载于*AAAI*秋季研讨会系列论文集：自主系统中的自信度。AAAI。 [20]

收到日期：2023年3月1日