Gemini - 通用智能体是否需要世界模型

2025年6月5日 2 分钟阅读

#world_model

> *本文探讨了通用智能体是否需要世界模型才能实现灵活的、面向目标的行为*

我为什么要阅读这篇文章?

该论文得出一个重要发现：**智能体本身就是世界模型。**

如何理解这个结论呢？通过论文证明，任何能够泛化到广泛简单目标导向任务的智能体，都必须已经学习了一个能够模拟其环境的预测模型。而且这个模型总是可以从该智能体中恢复出来的。

具体而言，论文证明了从任何满足在足够广泛的简单目标（比如将环境引导至期望状态）上具有遗憾界的目标条件策略中，都可以恢复出环境转移函数的有界误差近似值。

论文内容

这篇论文探讨了通用智能体是否需要世界模型才能实现灵活的、面向目标的行为。作者通过形式化的论证，证明了任何能够泛化到多步骤目标导向任务的智能体必须学习其环境的预测模型。该研究表明，随着智能体性能的提高或其能实现的目标复杂性的增加，所学到的世界模型也需要越来越精确。此外，该研究还提出了一种从智能体策略中提取世界模型的方法，并讨论了其对开发安全、通用智能体，以及限定智能体能力等方面的深远影响。

解决什么问题?

这篇论文旨在解决一个核心问题：**世界模型（world models）对于实现灵活、目标导向的行为是否是必需的，还是无模型学习（model-free learning）就已足够?** 换句话说，文章探讨了是否

目录

文章信息

字数

阅读时间

发布时间

更新时间

标签

#world_model

存在一条无需学习世界模型的"无模型捷径"来达到人类水平的AI，以及如果世界模型是必需的，那么它需要达到怎样的准确性和全面性才能支持给定的能力水平。论文提供了一个形式化的答案，证明了任何能够泛化到多步目标导向任务的智能体，都必须学习其环境的预测模型。

这不是一个新的问题

关于世界模型是否是实现人类水平AI的必要条件，长期以来一直存在争论，其核心在于学习模型的挑战与它们可能带来的潜在益处之间的权衡。例如，Brooks在"Intelligence without representation"一文中曾提出"**世界就是它最好的模型**"，智能行为无需学习显式的世界表示即可通过行动-感知循环在无模型智能体中涌现。然而，也有越来越多的证据表明，无模型智能体实际上可能学习了隐式世界模型，甚至可能学习了隐式规划算法。

论文要验证的科学假设

这篇文章要验证的科学假设是：**任何能够泛化到多步目标导向任务的智能体，都必须学习其环境的预测模型。**更具体地，论文证明，任何满足了针对足够多样化简单目标导向任务的遗憾（regret）界限的智能体，都必然学习了一个准确的环境预测模型。他们考虑的是在完全可观察的马尔可夫过程中，将通用智能体定义为能够针对大量简单目标导向任务（例如将环境引导至所需状态）满足遗憾界限的目标条件策略。然后，他们证明了对于任何这样的智能体，可以仅从其策略中恢复出环境转换函数（世界模型）的近似值，并且这种近似的误差会随着智能体性能的提高或其能实现的目标复杂性的增加而减小。

相关研究

论文在"相关工作"部分提到了多个相关研究领域：

> **逆强化学习（IRL）和逆规划（Inverse planning）**：这些领域研究如何从转换函数和最优策略中确定智能体的奖励函数或目标。该论文的成果补充了这些研究，它从智能体的策略和目标中恢复转换函数。
>
> 相关研究员/机构：Ng et al. (2000), Baker et al. (2007), Amin & Singh (2016)。
>
> **机械可解释性（Mechanistic Interpretability, MI）**：旨在揭示无模型智能体中的隐式世界模型。这通常涉及学习从策略网络的激活到表示状态的特征的映射。

相关研究员/机构：Abdou et al. (2021), Li et al. (2022), Gurnee & Tegmark (2023a,b), Karvonen (2024), Hou et al. (2023), Bush et al. (2025)。

因果世界模型（Causal world models）：Richens & Everitt (2024) 提出了一个类似的结果，表明能够适应足够大范围分布偏移的智能体必须学习因果世界模型。

相关研究员：Richens & Everitt (2024)。

LTL目标条件智能体（LTL goal-conditioned agents）：线性时序逻辑（LTL）是表达强化学习和规划中指令、目标和安全约束的自然选择。最近有多个实现了零样本泛化到任意LTL目标的智能体。

相关研究员/机构：Qiu et al. (2023, 2024), Jackermeier & Abate (2024, 2025), Vaezipoor et al. (2021), Kuo et al. (2020)。

表示定理（Representation theorems）：如Savage (1972) 和 Halpern & Piermont (2024) 的研究，证明满足某些理性公理的智能体行为，就如同它们正在最大化关于世界模型的效用函数的期望值。

相关研究员/机构：Savage (1972), Halpern & Piermont (2024), Von Neumann & Morgenstern (2007)。

良好调节器定理（Good regulator theorem）：该定理曾试图证明任何能够控制系统的智能体在某种意义上必须是该系统的模型。

相关研究员/机构：Conant & Ross Ashby (1970), Wentworth (2021)。

机构理论（Theories of agency）：在心理学和神经科学中，世界模型是几个重要理论的基础假设，例如建构主义感知理论、主动推理和意识理论。

相关研究员/机构：Gregory (1980), Friston (2010, 2013), Safron (2020)。

解决方案之关键

论文中解决方案的关键在于通过**证明归约(proof by reduction)**的方式来回答问题。具体来说：

假设智能体是“有界目标条件智能体”（bounded goal-conditioned agent）。这意味着智能体在某些有限深度n的目标导向任务上具有一定的（有下界的）能力。

推导出一个算法（Algorithm 1），该算法通过查询这个有界智能体的策略，来估计环境的转换函数（即世界模型）。该算法会向目标条件策略查询不同的复合目标（例如两种不兼容子目标之间的“二选一”决策）。

利用智能体的行动选择来推断世界模型：由于智能体满足遗憾界限，其行动选择编码了关于哪个子目标具有更高最大满足概率的信息，这些信息可以用来估计转换概率 Pss′(a)。

证明恢复的世界模型具有有界误差：论文进一步证明，该估计满足误差界限，即恢复的世界模型的准确性会随着智能体接近最优（δ → 0）或其能实现的任务序列深度n的增加而提高。 这个方法的关键在于，它表明**学习一个这样的目标条件策略在信息上等同于学习一个准确的世界模型。**

实验设计

论文通过实验验证了其从智能体中恢复世界模型的过程，以及模型准确性如何随着智能体泛化到更多任务（更长范围目标）而提高。实验设计如下：

环境：使用一个随机生成的、满足假设1的**受控马尔可夫过程（cMP）**，包含20个状态和5个动作，并具有稀疏的转换函数，以确保导航到给定目标状态不是一件微不足道的事。

智能体：采用**基于模型的智能体（model-based agent）**，其模型是通过从环境中随机策略下采样的不同长度（Nsamples ∈ {500, 1000, ..., 10,000}）的轨迹中学习的。重要的是，用于恢复世界模型的算法（Algorithm 2）不直接访问智能体的内部世界模型，而是仅将智能体的策略作为输入。

实验设置：

为每个样本量 Nsamples 训练10个智能体，每个智能体使用不同的随机种子生成经验轨迹。

对每个智能体，运行 Algorithm 2，测试不同的最大目标深度 N ∈ {10, 20, ..., 600}。

记录每个输入目标下的遗憾率 δ (1− P (τ |= ψn,m | π)/P (τ |= ψn,m | π∗))，并计算 Algorithm 2 返回的估计转换函数的所有状态-动作-结果元组的平均误差 ⟨ϵ⟩。

通过对给定智能体的 N（目标深度）与 ⟨δ⟩ 进行最小二乘回归，确定 Nmax(⟨δ⟩ = k)。

关键探索点：研究了即使智能体严重违反定义5的假设（例如，对某些深度n目标非常胜任，但对其他目标完全失败，导致遗憾界限被打破），算法能否仍然恢复转换函数。

定量评估的数据集和代码

用于定量评估的"数据集"实际上是**一个随机生成的受控马尔可夫过程（cMP）环境。**智能体通过从该环境中在随机策略下采样的轨迹进行训练，训练轨迹的长度 Nsamples 在500到10,000之间变化。

关于代码是否开源，**论文中没有明确提及代码。**

论文中的实验及结果

论文中的实验及结果**很好地支持了需要验证的科学假设。**

> **支持主要假设**：实验结果表明，从智能体中恢复的世界模型的平均误差 ⟨ϵ⟩ 随着智能体学习泛化到更高深度目标的能力（Nmax(⟨δ⟩ = 0.04)）而降低，其缩放比例约为 O(n−1/2)，与定理1中误差和遗憾之间的缩放关系一致。这表明，**当智能体能够处理更长的目标序列时，其隐式世界模型的准确性也会提高。**
>
> **鲁棒性验证**：即使智能体违反了定义5的严格假设（即在某些目标上达到最大遗憾率 δ = 1），Algorithm 2 仍能以较低的平均误差恢复转换函数。这表明，**即使智能体在某些目标上表现不佳，只要它在长时序目标上能达到较低的平均遗憾率，我们仍然可以准确地恢复其转换函数。**

论文的贡献

这篇论文的贡献是多方面的：

> **理论证明**：**提供了关于世界模型必要性的形式化证明。**它证明了任何能够泛化到足够广泛的简单、目标导向任务的智能体，都必然学习了一个准确的环境模型。这揭示了所有准确模拟环境所需的信息都包含在智能体的策略中。
>
> **驳斥“无模型捷径”**：**驳斥了“无模型捷径”实现通用AI的可能性。**如果智能体要泛化到长时序任务，学习世界模型是无法避免的。这重新激励了显式基于模型的架构研究。
>
> **解释涌现能力**：为基础模型中涌现能力提供了一个潜在机制的解释。为了在各种训练任务中最小化遗憾，智能体需要学习一个隐式世界模型，这反过来可以支持其泛化到从未明确训练过的任务。
>
> **AI安全的新视角**：提供了一个理论保证，即可以从任何足够强大的无模型智能体中提取准确的世界模型。模型的保真度会随着智能体能力的增强而提高，尤其是在智能体擅长实现长期目标时，这正是奖励欺骗等安全问题变得重要的领域。
>
> **对强AI的限制**：揭示了训练一个能够泛化到现实世界中各种任务的智能体是极其困难的，至少与学习一个准确的世界模型一样困难（甚至更难）。智能体的泛化能力最终受限于其理解世界运作方式的能力。
>
> **新算法**：在证明过程中，**推导出了从通用智能体中提取世界模型的新算法**（Algorithm 1和Algorithm 2），并在实验中展示了其有效性。

未来的工作

论文提出了一些未来工作的方向：

扩展目标类别：可以将分析扩展到超出定义3的不同目标类别，并识别出足以推断智能体已学习世界模型的“通用”任务集。这些任务可能对训练通用智能体有用。

世界模型提取算法的扩展：可以基于 Algorithm 1 开发更具可扩展性或适用于更通用环境的算法，并利用它们来提高智能体的安全性和可解释性。

结合机械可解释性：定理1为机械可解释性领域寻找隐式世界模型的工作提供了理论支持。未来的工作可以利用这种必要性来从世界模型的可学习性中推导出对智能体能力的新的基本界限。

部分可观察环境：需要探索在部分可观察环境中运行的智能体为了实现相同的行为灵活性，需要学习关于潜在变量的哪些信息。

心理学/神经科学理论的联系：结果为许多心理学和神经科学中的机构理论提供了强大的理论依据，证明了目标导向智能体必须获得世界模型才能实现一定程度的行为灵活性，无需先验地假设智能体具有世界模型。

参考

论文arxiv链接

作者x上的精彩介绍

我的论文NotebookLM Link