arXiv:2601.00696v1 [cs.LG] 2 Jan 2026

基于高维多模态观测的贝叶斯逆博弈

期刊标题 XX(X): 1–19©作者 2016。转载及授权事宜:
sagepub.co.uk/journalsPermissions.nav DOI:
10.1177/ToBeAssigned
www.sagepub.com/SAGE

Yash Jain[1*]、Xinjie Liu[1*]、Lasse Peters[2*]、David Fridovich-Keil[1], 和 Ufuk Topcu[1]

摘要
许多多智能体交互场景都可以自然地建模为非合作博弈，其中每个智能体的决策取决于其他智能体未来的行动。然而，要部署博弈论规划器进行自主决策，必须明确所有智能体的目标函数。为克服这一实际困难，近期研究开发了最大似然技术，用于求解逆博弈，并从交互数据中识别未知的智能体目标函数。遗憾的是，这些方法只能推断点估计，无法量化估计器的不确定性；因此，下游规划决策可能会过度自信地选择不安全动作。我们提出一种用于求解逆博弈问题的近似贝叶斯推断方法，该方法能够整合来自多种模态的观测数据，并在传感器观测有限的情况下，实时生成隐藏智能体目标函数的贝叶斯后验样本。具体而言，所提出的贝叶斯逆博弈框架在交互数据集上训练一个嵌入了可微纳什博弈求解器的结构化变分自编码器，且不需要智能体真实目标函数的标签。大量实验表明，我们的框架能够成功学习先验和后验分布，相较于基于最大似然估计的逆博弈方法提升推断质量，并在不牺牲效率的前提下实现更安全的下游决策。当轨迹信息缺乏信息量或不可用时，多模态推断还可通过利用额外的观测模态进一步降低不确定性。

关键词
概率推断、不确定性下的规划、机器人控制机器学习、博弈论、规划与学习一体化、用于控制的模型学习

1 引言

非合作博弈论（Bas¸ar 和 Olsder 1999）为建模多智能体交互式决策提供了一个具有严谨理论依据的框架，并已成为自主机器人进行交互感知运动规划的强大工具。现代数值方法（Fridovich-Keil 等 2020；Zhu 和 Borrelli，2023；Cleac'h 等人，2022；Li 等 2023a）显著加快了博弈论规划问题的求解，使其实时应用成为可能，并提升了其实用价值（Spica 等 2020；Wang 等，2021）。

在博弈论规划中，自车智能体（例如自主机器人）将自身及其他智能体的决策过程建模为目标函数可能相互冲突的耦合优化问题。当每个智能体的决策均为单方面最优时，联合决策即为纳什均衡。例如，图1 展示了交叉路口场景中的这种非合作交互：红色自车机器人与绿色对手（例如人类驾驶员）在避免碰撞的同时，都力求高效地达到各自的目标。

然而，在现实世界的部署中，这类方法通常以去中心化的方式使用，每个智能体只了解自身的目标函数；因此，为了计算考虑交互的纳什均衡运动规划，自车机器人必须推断未知的博弈参数，例如其他智能体的目标函数。在图1的例子中，为了安全交互并仍能高效到达目标，自车必须在瞬间根据观察到的对手行为，并在有条件时结合转向灯等视觉线索，推断对手是会直行还是左转。

从观测中推断未知博弈参数的问题称为逆博弈（Molloy 等，2022）。标准方法通常将最大似然估计（最大似然估计）问题（Peters 等，2023a；Liu 等人（2023）；Armstrong 等，2025；Mehr 等，2023；Hu 等，2025；Sun 等，2025）与一阶纳什均衡条件作为约束。然而，最大似然估计会产生固有的过度自信的点估计：它缺乏不确定性量化，并且经常忽略重要的先验知识，从而可能导致不安全的决策。例如，当对手刚刚进入交叉路口、尚未决定转弯操作时，最大似然估计方法可能会自信地断定对手将直行，而有经验的人类驾驶员则会保持谨慎，意识到对手可能激进左转（Liu 等，2024）。

*共同贡献[1]德克萨斯大学奥斯汀分校，美国德克萨斯州奥斯汀，邮编 78712[2]代尔夫特理工大学，荷兰代尔夫特，邮编 2600 AA

通讯作者： Xinjie Liu，德克萨斯大学奥斯汀分校，美国德克萨斯州奥斯汀，邮编 78712
电子邮箱：xinjie-liu@utexas.edu

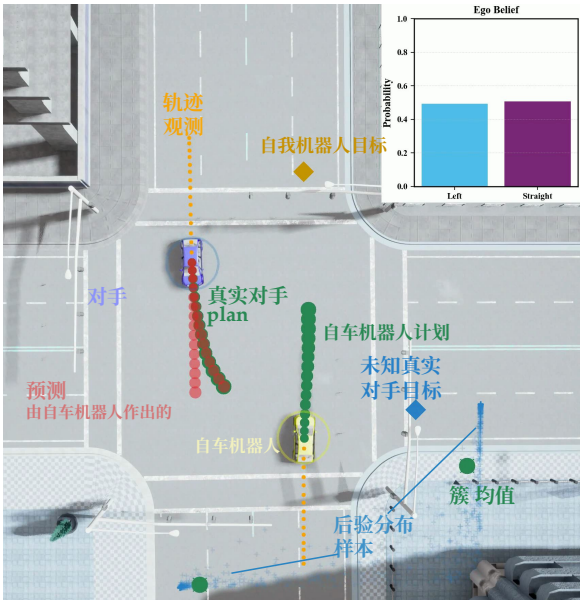


图1. 机器人与一位目标位置未知的对手驾驶员互动。我们将可微博弈求解器嵌入结构化变分自编码器中，以根据观测到的轨迹推断对手意图的分布，并结合对手的图像观测。

贝叶斯逆博弈则将参数推断转化为贝叶斯推断问题，该问题可以通过粒子滤波（Peters，2020）或无迹卡尔曼滤波器（Le Cleac'h 等人，2021）等在线滤波方法近似求解。最近，Liu 等人（2024）提出了一种结构化变分自编码器（变分自编码器）框架，通过在不含参数标签的交互轨迹数据集上进行摊销式离线训练，近似求解贝叶斯逆博弈问题。该生成式信念模型能够根据观测到的轨迹，推断非合作博弈中未知对手目标函数的后验分布，支持实时运行，并捕获多模态后验分布，从而提供不确定性量化，提升下游运动规划的安全性。

然而，先前的逆博弈方法几乎完全基于低维观测，即状态轨迹。在现实世界的交互中，其他观测模态——例如视觉线索（转向灯、车辆类型等）——可能包含有关对手意图的关键信息，尤其是在新智能体刚进入场景且几乎没有或完全没有轨迹历史可用时。在高度动态的交互环境中，利用纯粹的行为观测之外的上下文信息的能力可能关乎安全性，这促使人们探索基于多种传感器的高维多模态观测进行逆博弈的方法。

为此，本文对Liu 等人（2024）（即将刊出；将发表于机器人学算法基础 *XVI*，施普林格·自然），并提出了一种易于处理的贝叶斯逆博弈框架，将可微纳什博弈求解器嵌入变分自编码器。所提出的框架：

- **（可处理性与可解释性）**利用可微纳什博弈求解器提供的解析梯度，学习结构化生成式信念模型，从而使多模态后验推断变得可处理，同时为下游运动计划器保留可解释性；
- **（多模态观测融合）**将原始图像中的视觉线索与智能体交互的部分状态观测相融合，以进行摊销贝叶斯推断；当轨迹数据不可用或信息量不足时，可降低不确定性；
- **（下游决策）**在实时运行的同时，相较于基于最大似然估计的逆博弈方法，实现更安全、更高效的不确定性感知决策。

除会议版本之外，本文还 :

- 将原有的贝叶斯逆博弈框架及其架构推广至多种观测模态，包括高维原始图像；
- 在 CARLA 仿真中开展了新的实验研究（Dosovitskiy 等，2017），采用逼真的视觉观测和低层跟踪控制器；
- 系统评估了多模态观测对贝叶斯推断质量及下游运动规划性能的影响。

在本文中，“多模态”既可指（i）具有多个峰值的先验分布/后验分布，也可指（ii）多种观测模态（例如轨迹和图像）；具体含义可从上下文中清楚判断。

2 相关工作

本节概述动态博弈论相关文献，重点介绍正向博弈（第 2.1 节）和逆博弈（第 2.2 节）。*

2.1 动态博弈

本文关注非合作动态博弈，其中智能体可以具有任意且可能相互冲突的目标函数，不进行串谋，并随时间推移依次做出决策（Bas¸ar 和 Olsder 1999）。由于我们假设智能体同时采取行动，且不存在领导者—跟随者层级，并考虑智能体的决策通过目标函数和约束相互耦合，因此本文聚焦于广义纳什均衡问题（GNEPs）；近年来，针对这类问题已出现多种高效的计算方法。

由于在Laine 等，2023反馈信息结构下求解此类问题面临计算上的挑战（Bas¸ar 和 Olsder 1999），大多数研究转而寻求开环纳什均衡，其中参与者一次性选择完整的动作序列，即开环策略。大量研究采用迭代最佳响应算法，通过反复求解单智能体优化问题来求得开环纳什均衡（Williams 等，2018；Spica 等，2020；Wang 等，2019；Spica 等，2020；Wang 等，2021；Schwarting 等，2019）。近年来，基于序贯

*第 2.1 节和第 2.2 节大部分内容改编自我们先前发表的会议论文（Liu 等，2024）。

二次近似的方法已被提出（Cleac'h 等人，2022；Zhu 和 Borrelli 2023），旨在通过在每次迭代中同时更新所有参与者的开环策略来加快收敛。最后，由于开环广义纳什均衡问题的一阶必要条件可表示为混合互补问题（MCP）（Facchinei 和 Pang，2003），已有若干研究（Liu 等人（2023）；Peters 等人，2024）使用成熟的混合互补问题求解器求解广义纳什均衡（GNE）（Billups 等人，1997；Dirkse 和 Ferris（1995））。本研究建立在后一种方法之上。

2.2 逆动态博弈

逆博弈研究如何根据智能体行为的观测来推断未知博弈参数，例如目标函数参数（Waugh 等人，2011）。近年来，若干方法将单智能体逆最优控制（IOC）和逆强化学习（IRL）技术扩展到多智能体交互场景。例如，（Rothfuß 等人，2017；Awasthi 和 Lamperski，2020）中的方法，在给定完整状态-控制观测的情况下，通过最小化智能体一阶必要条件的残差来推断未知目标参数。这种方法在（Inga 等人，2019）中进一步扩展到了最大熵场景。

近期研究（Peters 等，2023b）提出，在将开环纳什均衡（OLNE）的 Karush–Kuhn–Tucker（KKT）条件作为约束的同时，最大化观测似然。该方法仅需部分状态观测，并且能够处理受噪声污染的数据。Li 等（2023b）和Liu 等人（2023年）提出将最大似然估计方法（Peters 等，2023b）扩展到带不等式约束的逆反馈博弈和开环博弈，并利用广义纳什均衡相对于问题参数的方向可微性。为摊销最大似然估计的计算，Geiger 和 Straehle （2021）以及Liu 等人（2023年）展示了与神经网络（NN）组件的集成。

一般而言，最大似然估计解可视为贝叶斯后验分布的点估计，前提是采用均匀先验（Murphy2022，第 4 章）。当多个参数值都能同样好地解释观测时，这种简化可能导致病态问题，使 MLE 逆博弈恢复出可能不准确的估计（Li 等人，2023b）。此外，在运动规划中，使用未考虑不确定性的点估计可能导致不安全规划（Peters 等人，2024；Hu 和 Fisac，2022）。

为了解决这些问题，一些研究采用逆博弈的贝叶斯视角（Le Cleac'h 等人，2021；Peters，2020），旨在推断隐藏博弈参数的后验分布，同时纳入先验知识。由于这些问题中的精确贝叶斯推断难以处理，信念更新可以通过粒子滤波器近似（Peters，2020）。然而，这种方法需要在线求解大量均衡问题来维持信念分布，带来了显著的计算负担。西格玛点近似（Le Cleac'h 等人，2021）减少了所需样本的数量，但将估计器限制在单峰不确定性模型中。

为了高效地获得多假设预测，Diehl 等人（2023年）和 Lidard 等人 （2023年）将博弈论层融入用于运动预测的神经网络中。两种方法均表明，博弈的归纳偏置能够提升在真实世界人体数据集上的性能。在（Lidard 等人，2023年）中， Diehl 等人 （2023年）将势博弈（Monderer 和 Shapley，1996）的解嵌入神经网络，以预测轨迹候选项。然而，这两种方法都仅限于预测固定数量的轨迹，且对所学习模型缺乏明确的贝叶斯解释。

为克服最大似然估计方法的局限性，同时避免对连续博弈参数分布进行精确贝叶斯推断时的难以处理性，我们提出通过变分自编码器近似后验分布（Kingma 和 Welling 2014），并在训练过程中嵌入可微博弈求解器（Liu 等人（2023））。所提出的方法可利用观测到的交互构成的无标签数据集进行训练，能够自然地处理连续、多模态的参数分布和轨迹分布，并且在运行时从后验分布中采样参数时无需计算博弈解。此外，为了在轨迹观测信息不足或不可用时利用视觉线索中关键但细微的上下文信息，我们将所提出的框架扩展为联合处理多模态观测。

2.3 多模态变分自编码器

在机器学习文献中，已有许多多模态变分自编码器框架被开发出来，用于融合不同模态（例如视觉、文本和音频）中的信息，尤其注重处理缺失模态并实现跨模态生成（Schonfeld et al. 2019；Shiet al.2019；Wu and Goodman2018；Suzuki etal. 2017；Sutter et al. 2021）。

在机器人学中，多模态变分自编码器常用于传感器融合，即使某些模态不可用，也能基于异构输入支持感知、预测和控制（Zhou 等 2021；Zambelli 等 2020；Meo 和 Lanillos 2021）。例如，Zhou 等 (2021) 采用变分专家乘积形式融合 RGB 图像和深度图像，以实现定位。Zambelli 等 (2020) 从状态、视觉、触觉和声音中学习共享感觉运动潜在空间，用于人形机器人控制。Meo 和 Lanillos(2021) 将多模态变分自编码器融入主动推理控制器，使其能够根据原始图像和本体感受状态估计，对工业机械臂进行反应式控制。

在自动驾驶中，多模态变分自编码器主要用于运动预测（Lee etal. 2017；Salzmann etal. 2020；Yuan et al. 2021），通常以智能体的过去状态以及场景图或语义地图等上下文线索作为条件。与那些明确要求不同模态共享潜在空间的多模态变分自编码器方法（Schonfeld et al.2019）不同，这些预测模型通常将多模态输入视为条件变量，而不是通过联合潜在表示对其进行对齐。只有在测试时可能缺失某些模态的情况下（例如，新智能体进入场景时，我们可能只有摄像机图像，却没有观察到其行为历史），学习共享潜在空间才最为有益（Schonfeld etal.2019）。

更广泛而言，物理信息变分自编码器通过仿真器、约束层或基于物理的先验分布／动力学引入可微的物理结构，使潜在空间的部分内容扎根于具有物理意义的量，同时保留标准的摊销变分推断（Zhongand Meidani2023；Thoreau et al. 2025；Takeishi andKalousis2021）。沿着这一思路，我们将可微纳什博弈求解器嵌入变分自编码器，使采样以可解释的博弈参数为基础，同时学习跨模态的联合潜在空间，即使智能体的历史状态有所缺失，该空间仍能保持有效。

3 预备知识

本文从参数化广义纳什均衡问题（*GNEP*）的视角，研究理性且以自身利益为重的智能体之间的交互。在这一设定中，N 个智能体中的每一个都旨在独立地最小化自身成本，并充分意识到其余智能体——即其“对手”——同样在追求各自的个体目标。我们通过 N 个相互关联的约束优化问题来构建一个参数化的 N智能体 GNEP，表示为

$$\mathcal{S}_\theta^i(\tau^{\neg i}) := \arg\min_{\tau_i} \quad J_\theta^i(\tau^i, \tau^{\neg i}) \tag{1a}$$

$$\text{s.t.} \quad g_\theta^i(\tau^i, \tau^{\neg i}) \geq 0, \tag{1b}$$

其中，$\theta \in \mathbb{R}^p$ 对智能体的目标函数和约束进行参数化。对于每个智能体 $i \in [N] := \{1, \ldots, N\}$，$J_\theta^i$表示其成本函数，$g_\theta^i$ 表示其私有约束。值得注意的是，智能体 i 的成本和约束不仅取决于其自身策略$\tau^i \in \mathbb{R}^{m_i}$，还取决于其他智能体的集体策略 $\tau^{\neg i} \in \mathbb{R}^{\sum_{j\in[N]\setminus\{i\}} m_j}$。本文中，我们交替使用智能体和博弈参与者这两个术语，指代博弈中的 N 个决策实体之一。

广义纳什均衡。给定参数 θ，GNEP 的解是一个广义纳什均衡：一个策略组合 $\tau^* := (\tau^{1*}, \ldots, \tau^{N*})$，其中每个智能体的策略都是对其他所有智能体策略的最优响应，这意味着

$$\tau^{i*} \in \mathcal{S}_\theta^i(\tau^{\neg i*}), \quad \forall i \in [N]. \tag{2}$$

本质上，在广义纳什均衡中，没有任何智能体能够通过单方面偏离到另一种可行策略来降低自己的成本。

示例：博弈论运动规划。前述博弈论框架可以自然地用于机器人与其他不受控制的智能体（人类或其他机器人）交互时的在线轨迹规划。图1展示了这样一种场景。在此，每个智能体 i的策略τ_i 对应一条轨迹（由状态和控制量构成的有限时域序列）。从实践者的角度来看，设置博弈论运动规划的过程与设置标准基于优化的运动规划类似：实践者编写代码，为每个智能体 i指定约束 $g_\theta^i(\cdot)$，以编码状态/输入限制、动力学和障碍物规避，并指定目标函数 $J_\theta^i(\cdot)$ 来体现 i^{th} 智能体的偏好，例如朝目标位置推进。通常，约束和目标函数都取决于机器人及其对手的当前状态（例如，适应变化的车道边界或速度限制，并避免碰撞）。

在线执行期间，机器人调用博弈求解器，以滚动时域方式求解该轨迹博弈的均衡解，从而形成一种类似于模型预测控制（MPC）的决策范式：模型预测博弈（*MPGP*）。在此背景下，博弈的均衡解有两个作用：按照索引 1 表示自车智能体的约定，均衡解中的对手轨迹 $\tau^{\neg 1*}$ 提供了对手未来决策的博弈论预测，而自车智能体的轨迹 τ^{1*} 则给出机器人相应的最佳响应。

备注 1。方程 (2)的结构——每个策略 τ^{i*} 都取决于所有其他参与者的策略$\tau^{\neg i*}$，反之亦然——凸显了智能体策略之间的相互依赖性。参与者不能仅仅预测其他人的策略，再据此优化自己的策略，因为改变自身策略反过来也可能影响这些预测。因此，求解广义纳什均衡问题需要同时进行预测与规划方面的推理。这一特性使广义纳什均衡问题能够有效建模紧密耦合、相互依存的决策。

博弈参数的作用 θ。多智能体轨迹博弈与单智能体轨迹优化的一个主要区别，在于需要为参与其中的每一个智能体指定成本和约束。在现实场景中，机器人往往无法完全了解对手的目标函数、动力学或状态，因此难以构建一个定义完备的博弈论模型。这一点促使我们采用前文介绍的参数化博弈形式：在本文后续内容中，θ 将表示博弈中的不确定因素。在博弈论运动规划的场景中，θ 通常涵盖对手成本函数和约束的特征，例如其未明确指定的目标车道或期望速度。为简化符号表示，我们将博弈（1）简洁地表示为问题组成部分的参数化集合 $\Gamma(\theta) := (\{J_\theta^i, g_\theta^i\}_{i\in[N]})$。在接下来的章节中，我们将讨论如何根据观测行为在线估计这些参数。

4 方法

机器人与他人互动时，通常无法事先知晓对方的意图。因此，为了安全且高效地进行互动，机器人必须根据观测到的参与者行为在线推断他人的意图。我们的主要贡献正是从博弈论视角解决这一问题。

我们的方法的核心思想，是对第3节提出的问题进行反向求解：假设观测到的行为是通过求解一个参数未知的博弈而产生的，我们便尝试找出能够解释该行为的博弈参数。因此，这一过程通常称为逆博弈问题（Waugh 等，2011）。

以往关于逆博弈的研究，例如 Peters 等 (2023b)，Li 等 (2023b和 Clarke 等(2023年)，主要侧重于通过最大化观测行为的似然来推断博弈参数。尽管这种最大似然估计方法被广泛采用，但它存在明显的局限性：

(i) 它只能给出 θ 的单一“最佳拟合”估计，因此难以捕捉或量化不确定性，而这对运动规划等任务至关重要；(ii) 当可用观测提供的信息不足或含糊不清时，该方法可能失效，导致参数恢复效果不佳。我们通过第 5 节中的实验展示了这些局限性。

此外，与此前关于逆博弈的研究（包括本研究所依据的会议论文（Liu 等，2024））不同，本文提出的方法不仅纳入低维状态观测（如位置和速度），还纳入高维传感器数据（如图像）。正如我们在第 5 节所展示的，这种在逆博弈中处理视觉线索的新能力，使我们的方法能够比以往方法更快地降低对其他参与者意图的不确定性，从而实现更安全、更高效的交互。

4.1 逆博弈的贝叶斯视角

我们采用如下的贝叶斯视角来研究逆博弈。设 $y = (y_{\text{traj}},\ y_{\text{img}}) \in \mathbb{R}^{d_{y_{\text{traj}}}} \times \mathbb{R}^{d_{y_{\text{img}}}}$ 表示与参数为 θ 的博弈之解相对应的联合观测，该观测由两种模态构成：轨迹观测 y_{traj} 和图像观测 y_{img}。

我们的目标是根据这一多模态观测计算博弈参数 θ 的后验分布（信念），即，

$$b(\theta) := p(\theta \mid y) \tag{3}$$

与仅得出点估计 $\theta_{\text{MLE}} \in \operatorname{argmax}_\theta p(y \mid \theta)$ 的最大似然估计公式不同，这种贝叶斯方法能够推断未知博弈参数 θ 的完整后验分布，并自然地纳入先验信息。

4.1.1 观测模态 轨迹数据和图像数据具有不同的特征，下面将对此进行描述。

轨迹观测， y_{traj}。轨迹观测是低维的部分状态观测，涵盖近期过去的一个固定时滞窗口（例如，在我们的示例中，是最近 10 个时间步的含噪位置和速度测量）。这种模态让机器人感知物理状态以及其他参与者最近采取的行动。采用逆博弈方法时，我们假设轨迹观测是所有智能体在参数未知的博弈中进行博弈所直接产生的结果。与这一假设一致，和以往的研究（Le Cleac’h 等人，2021；Li 等人，2023b；Peters 等，2023b）一样，我们为轨迹观测选择高斯观测模型，即

$$p(y_{\text{traj}} \mid \theta) := \mathcal{N}(\underbrace{(h_{\text{traj}} \circ \mathcal{T}_\Gamma)(\theta)}_{\mu_{y_{\text{traj}}}(\theta)}, \Sigma_{y_{\text{traj}}}), \tag{4}$$

其中，$\mathcal{T}_\Gamma$ 表示一个博弈求解器，它将博弈参数 θ 映射为博弈 $\Gamma(\theta)$的均衡轨迹剖面 τ^*；h_{traj} 表示一个函数，它将该均衡轨迹剖面 τ^* 映射为部分轨迹观测分布的均值（例如，从轨迹中提取位置和速度）。为简单起见，我们假设协方差矩阵 $\Sigma_{y_{\text{traj}}}$ 固定且已知。

图像观测， y_{img}。图像观测能提供关于其他参与者意图的额外视觉线索。不同于轨迹观测，图像观测难以仅通过简单的高斯观测模型与博弈解建立关联。例如，在图1中的交叉路口场景里，相同的轨迹轮廓 τ^* 可能对应截然不同的图像：场景可能发生在不同的天气状况和一天中的时段，涉及不同类型、不同颜色的汽车，且道路几何形状也可能不同。然而，驾驶行为与图像观测之间存在微妙的相关性：例如，跑车可能更倾向于激进驾驶，卡车可能较少突然转弯，并倾向于留在右侧车道；而在雨天，驾驶员可能会更加谨慎。

通过简单的、由用户指定的观测模型来手动建模这些细微的相关性是难以实现的。相反，我们将直接从包含联合观测 $\mathcal{D} = \{y_k = (y_{\text{traj}_k},\ y_{\text{img}_k}) \mid y_k \sim p(y)\ \forall k \in [K]\}$ 的无标签数据集中学习隐式观测模型。

4.2 自动编码贝叶斯逆博弈

理论上，公式（3）中的逆博弈问题可以通过贝叶斯定则简单求解

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}.$$

然而，除了复杂传感器数据（如图像）通常没有可用观测模型这一难题之外，还有若干挑战使得直接推断后验分布难以实现：

1. 先验分布 $p(\theta)$通常不可用，因此必须从数据中学习。2. 归一化常数 $p(y) = \int p(y \mid \theta)p(\theta)\mathrm{d}\theta$ 的计算在实践中难以实现，因为需要对 θ 进行边缘化。3. 先验分布 $p(\theta)$ 和后验分布 $p(\theta \mid y)$ 通常都是非高斯的，甚至可能是多模态的，因此难以用其概率密度函数（PDF）显式表示。

先前的工作（Le Cleac’h 等人，2021）通过使用无迹卡尔曼滤波器（UKF）进行近似贝叶斯推断，部分缓解了这些挑战；但该方法仅适用于单峰不确定性模型和低维观测，而且每次信念更新都需要求解多个博弈，因此带来了计算方面的挑战。

幸运的是，正如我们在第 5 节中所展示的，逆博弈的许多实际应用并不需要显式计算信念的概率密度函数，$b(\theta)$。相反，信念的 $\theta \sim b(\theta)$生成模型——即能够让我们抽取样本的模型——通常就已足够。本节将展示如何通过构建博弈论变分自编码器（VAE），从一个包含 $\mathcal{D} = \{y_k = (y_{\text{traj}_k}, y_{\text{img}_k}) \mid y_k \sim p(y)\ \forall k \in [K]\}$ 观测到的交互的无标签数据集，学习这样的生成模型。

4.2.1 引入潜变量模型为了学习信念的生成模型 $b(\theta)$，我们引入一个潜变量模型，如图2所示。遵循这一模型结构，除 $p(y_{\text{traj}} \mid \theta)$ 的轨迹观测模型外，公式（4），我们还定义以下

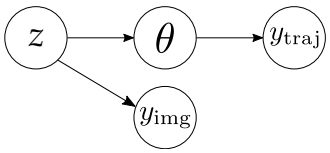


图 2. 我们将观测图像 y_{img} 和轨迹 y_{traj} 通过潜变量 z与博弈参数 θ 联系起来的潜变量模型的贝叶斯网络表示。请注意，根据构造，给定潜变量 z后，观测图像 y_{img} 和轨迹 y_{traj} 在统计上相关，但条件独立。

分布：

$$p(z) \coloneqq \mathcal{N}(0, I), \tag{5a}$$

$$p_\phi(\theta \mid z) \coloneqq \delta(\theta - d_\phi^\Gamma(z)), \tag{5b}$$

$$p_\phi(y_{\text{img}} \mid z) \coloneqq \mathcal{N}(d_\phi^{\text{img}}(z), \Sigma_{y_{\text{img}}}), \tag{5c}$$

其中，δ 表示狄拉克δ函数，d_ϕ^{img} 和 d_ϕ^Γ分别表示将潜变量 $z \in \mathbb{R}^{d_z}$映射到博弈参数 θ 和图像观测 $\mu_{y_{\text{img}}}$的均值的解码器神经网络，而 $\Sigma_{y_{\text{img}}}$ 表示图像观测的协方差矩阵。

由于 θ和 z之间存在确定性关系，我们可以进一步将轨迹观测模型直接写成以 z 为条件的形式：

$$p_\phi(y_{\text{traj}} \mid z) \coloneqq \mathcal{N}((h_{\text{traj}} \circ \mathcal{T}_\Gamma \circ d_\phi^\Gamma)(z), \Sigma_{y_{\text{traj}}}), \tag{6}$$

其中，$\Sigma_{y_{\text{traj}}}$ 表示轨迹观测的协方差矩阵。

自然，只有在能够解释数据时，潜变量模型才有用。为此，我们将寻求一种参数化方式 ϕ^*，使得由此诱导的数据分布

$$p_{\phi^*}(\overbrace{y_{\text{traj}}, y_{\text{img}}}^{y}) \coloneqq \int p_{\phi^*}(y_{\text{traj}} \mid z) p_{\phi^*}(y_{\text{img}} \mid z) p(z) \mathrm{d}z,$$

与真实数据分布$p(y)$相匹配，即$D_{\text{KL}}(p_\phi(y) \parallel p(y)) \approx 0$，其中

$$D_{\text{KL}}(p \parallel q) \coloneqq \mathbb{E}_{x \sim p(x)} [\log p(x) - \log q(x)]$$

表示任意分布 p和 q之间的 Kullback–Leibler（KL）散度。

4.2.2 潜变量模型的性质新引入的潜变量 z 没有任何物理意义。然而，所选的模型结构具有一项实用性质：它使我们能够将博弈参数的先验分布和后验分布表示为关于潜变量 z的推断问题。具体而言，我们可以将博弈参数的先验分布表示为

$$p_\phi(\theta) \coloneqq \int p_\phi(\theta \mid z) p(z) \mathrm{d}z, \tag{7}$$

并将博弈参数的后验分布表示为[†]

$$p_\phi(\theta \mid y) \coloneqq \int p_\phi(\theta \mid z) p_\phi(z \mid y) \mathrm{d}z. \tag{8}$$

这一分析表明，我们可以通过从 $p(z)$ 中采样（根据构造，该分布已知为多元高斯分布），再将样本输入解码器网络 d_ϕ^Γ，从而生成来自 $d_\phi^\Gamma(p_\phi)$ 的样本。类似地，如果将来自 $p_\phi(z \mid y)$ 的样本输入解码器网络d_ϕ^Γ，就能生成来自 $p_\phi(\theta \mid y)$ 的样本。

然而，直接从精确后验分布 $p_\phi(z \mid y)$中采样，需要对 θ 和 z进行边缘化，这通常难以处理。因此，我们将拟合一个高斯替代模型 $q_\psi(z \mid y)$，使得 $D_{\text{KL}}(q_\psi(z \mid y) \parallel p_\phi(z \mid y)) \approx 0$，这与摊销变分推断中的常见做法一致（Murphy 2023年）。由于 d_ϕ^Γ 高度非线性，尽管在潜在空间中采用了高斯近似，这仍能使我们捕捉博弈参数的非高斯、多模态分布 θ。我们将在第 5 节中展示我们的方法捕捉此类多模态分布的能力。

在讨论如何训练潜变量模型 p_ϕ 和替代模型 q_ψ之前，我们先简要说明一下我们的方法与变分自编码器术语之间的联系。

4.2.3 与变分自编码器的联系在变分自编码器的术语中，替代模型 $q_\psi(z \mid y)$ 称为编码器。此外，分布

$$p_\phi(\overbrace{y_{\text{traj}}, y_{\text{img}}}^{y} \mid z) \coloneqq p(y_{\text{traj}} \mid z) p(y_{\text{img}} \mid z) \tag{9}$$

在变分自编码器的术语中对应解码器 。

我们的方法与传统变分自编码器之间的关键区别，在于该解码器具有特殊的结构。传统变分自编码器采用无结构的神经网络作为观测模型，而我们的观测模型则将博弈求解器 $\mathcal{T}_\Gamma$ 作为一个特殊的“层”。解码流程中的这一博弈论“层”赋予了解码器神经网络输出可解释的结构 d_ϕ^Γ，促使其预测隐藏博弈参数 θ。图 3展示了如何将我们的方法理解为结构化变分自编码器。

4.3 训练博弈论引导的变分自编码器

下面，我们概述联合优化潜变量模型的模型参数 p_ϕ 和编码器 q_ψ 的过程。在明确其与传统变分自编码器的联系后，我们首先回顾高层次的训练过程，该过程与文献（Kingma 和 Welling 2014；Murphy 2023年）中的方法密切一致，详见第 4.3.1 节。接下来，第 4.3.2 节将重点介绍由于嵌入博弈求解器而产生的特殊注意事项 $\mathcal{T}_\Gamma$。

4.3.1 高层次训练过程 潜变量模型 p_ϕ 的参数和编码器 q_ψ 可以通过在数据分布 $p(y)$上最大化证据下界（ELBO）来联合优化。具体而言，我们可以表示为

[†]在公式（8）中，θ 和 y 在给定 z 的条件下具有条件独立性，这是由于 θ 和 z之间存在确定性关系，这意味着$p_\phi(\theta \mid y, z) \equiv p_\phi(\theta \mid z)$。

联合观测的对数似然 y 为

$$\begin{aligned}
&\log p_\phi(y) \\
&= \log \int p_\phi(y \mid z) p(z) \mathrm{d}z && \text{(10a)} \\
&= \log \int \frac{p_\phi(y \mid z) p(z)}{q_\psi(z \mid y)} q_\psi(z \mid y) \mathrm{d}z && \text{(10b)} \\
&= \log \mathbb{E}_{z \sim q_\psi(z \mid y)} \left[\frac{p_\phi(y \mid z) p(z)}{q_\psi(z \mid y)} \right] && \text{(10c)} \\
&\geq \mathbb{E}_{z \sim q_\psi(z \mid y)} \left[\log p_\phi(y \mid z) + \log p(z) - \log q_\psi(z \mid y) \right] \\
&=: \ell(\phi, \psi, y). && \text{(10d)}
\end{aligned}$$

因此，为了拟合参数 ϕ 和 ψ，我们可以最大化[‡]

$$\phi^*, \psi^* \in \arg\max_{\phi, \psi} \mathbb{E}_{y \sim p(y)} \left[\ell(\phi, \psi, y) \right]. \tag{11}$$

实际上，这一优化问题通过对数据集 $\mathcal{D}$ 中的（批量）样本进行随机梯度上升来求解，并以样本平均替代外层期望。这要求通过涉及观测模型 $p_\phi(y \mid z)$的内层期望 $q_\psi(z \mid y)$传播梯度。

与传统变分自编码器类似，我们可以使用“重参数化技巧”（Kingma 和 Welling 2014）来计算内层期望的梯度。然而，即使采用这一技巧，计算梯度$\nabla_{\phi,\ \psi}\ell$仍面临额外挑战：梯度需要通过解码器 $p_\phi(y \mid z)$ 在方程（9）中的轨迹观测模型 $p_\phi(y_{\text{traj}} \mid z)$传播，而这又要求梯度通过方程（6）中的博弈求解器 $\mathcal{T}_\Gamma$ 在方程（6）中传播。我们将在下一节讨论如何应对这一挑战。

4.3.2 通过博弈论知情观测模型传播梯度 为了优化 方程（11），我们需要计算解码器 $p_\phi(y \mid z)$ 相对于解码器参数 ϕ和潜变量样本 z的梯度。尽管这一过程中的大部分步骤都可以通过自动微分完成，但展开方程（6）中的链式法则可知，这两种操作都涉及通过博弈求解器 $\mathcal{T}_\Gamma$ 进行反向传播。为了使这一反向传播可行，我们采用了先前工作中提出的隐式微分方法（Liu 等人（2023））。为完整起见，我们在此简要概述该方法。

局部均衡条件。为了便于梯度通过博弈求解器传播，我们分析博弈的局部均衡条件。对于 $\Gamma(\theta)$中定义的博弈方程（1），这些局部均衡条件对应于所有参与者的 Karush–Kuhn–Tucker（KKT）条件堆叠：

$$\forall i \in [N] : \begin{cases} \nabla_{\tau^i} \mathcal{L}^i(\tau^i, \tau^{\neg i}, \lambda^i, \theta) = 0, \\ 0 \leq g_\theta^i(\tau^i, \tau^{\neg i}) \perp \lambda^i \geq 0, \end{cases} \tag{12}$$

其中 $\mathcal{L}^i$ 表示博弈参与者 i的拉格朗日函数，

$$\mathcal{L}^i(\tau^i, \tau^{\neg i}, \lambda^i, \theta) := J_\theta^i(\tau^i, \tau^{\neg i}, \theta) - \lambda^{i\top} g_\theta^i(\tau^i, \tau^{\neg i}, \theta).$$

我们使用 PATH 求解器（Dirkse 和 Ferris（1995））求出满足这些条件的原变量 τ^{i*} 和对偶变量 λ^{i*}。

隐式微分。 KKT 条件在原变量解 τ^{i*}、对偶变量 λ^{i*} 和博弈参数 θ 之间建立了隐式关系。我们利用这一关系，按如下方式计算解 τ^{i*} 对博弈参数 θ 的敏感度。令 $\mathcal{I}^i$表示在解处取等号的 $g_\theta^i(\tau^{i*}、\ \tau^{\neg i*}、\ \theta)$ 的行，令 $\neg\mathcal{I}^i$ 表示该集合的补集，令 $[\cdot]_\mathcal{I}$ 表示返回与索引 $\mathcal{I}$ 关联行的算子。采用这一记号，我们定义 F_Γ 以提取在解处取等号的所有条件，

$$F_\Gamma(\tau^*, \lambda^*, \theta) := \begin{bmatrix} \vdots \\ \nabla_{\tau^i} \mathcal{L}^i(\tau^i, \tau^{\neg i}, \lambda^i, \theta) \\ \left[g_\theta^i(\tau^i, \tau^{\neg i}, \theta) \right]_{\mathcal{I}^i} \\ \left[\lambda^i \right]_{\neg \mathcal{I}^i} \\ \vdots \end{bmatrix},$$

针对所有参与者 $i \in [N]$堆叠。除此之外，我们定义 v^*:=(τ^1，λ^1，……，τ^N，λ^N)为所有参与者的原变量和对偶变量组成的堆叠向量。假设没有任何约束处于弱活跃状态（即原变量约束取等号且对应的对偶变量为零），那么对 θ 的微小变化必须由原变量和对偶变量的变化来补偿，使得F_Γ 仍为零。从数学上讲，这意味着全导数 F_Γ 关于博弈参数 θ 必须为零。展开这个全导数可得

$$\frac{\mathrm{d}F_\Gamma}{\mathrm{d}\theta} = \frac{\partial F_\Gamma}{\partial \theta} + \frac{\partial F_\Gamma}{\partial v^*} \frac{\partial v^*}{\partial \theta} = 0, \tag{13}$$

这为我们提供了一个线性方程组，我们可以通过求解该方程组得到 $\frac{\partial v^*}{\partial \theta}$，从而揭示博弈解关于博弈参数的偏导数。[§]

备注 2。 请注意，与博弈求解器的前向传播相比，方程 (13) 中的导数计算开销很低：由于我们已在前向传播中计算了 v^*，唯一需要额外进行的计算是对活跃条件关于 θ 和 v^* 求偏导，以建立线性方程组方程 (13)。我们的开源实现（*JGTP 团队，2023年*）通过与 *Julia* 中的自动微分框架紧密集成，实现了这一流程的自动化。

4.3.3 关于推理时间的说明。在推理时，为了从估计后验分布中生成样本$q_{\psi,\ \phi}(\theta \mid y)$: $=\int p_\phi(\theta \mid z) q_\psi(z \mid y) \mathrm{d}z$，我们无需评估博弈求解器 $\mathcal{T}_\Gamma$ 或图像解码器 d_ϕ^{img}：后验采样只涉及对神经网络 d_ϕ^Γ 和 e_ψ的评估；参见图 3。因此，给定由图像和轨迹构成的联合观测 y，我们可以运行编码器一次来

[‡]很容易验证，该目标函数的极大化点也会使数据的对数似然最大化，同时使真实后验分布与替代后验分布之间的 KL 散度最小化，因为 ℓ 可以等价地写为$\ell(\phi, \psi, y) = \log p_\phi(y) - D\mathrm{KL}\left(q_\psi(z \mid y) \parallel p_\phi(z \mid y)\right)$，参见 Murphy（2023年）。[§]此处不讨论某些边界情况。例如，解可能包含弱活跃约束；或者，当 $\frac{\partial v^*}{\partial \theta}$ 为秩亏时，方程（13）可能不存在唯一解。关于如何处理这些边界情况，请参阅我们先前的工作（Liu 等人（2023））。

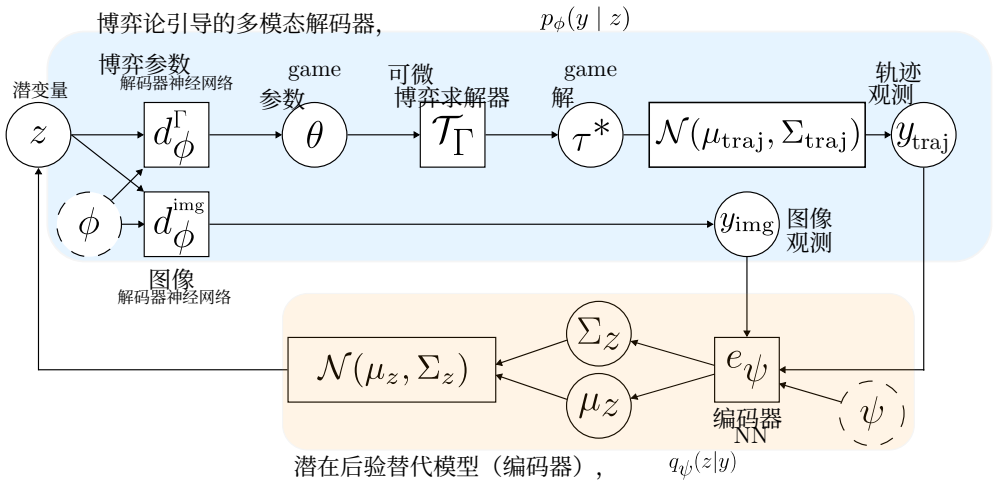


图 3. 用于生成式贝叶斯逆博弈的结构化变分自编码器概览。

获得潜变量分布 $q_\psi(z \mid y)$，然后反复从该分布中采样，以较低的计算成本从 $q_{\psi,\ \phi}(\theta \mid y)$中恢复大量后验样本。因此，该推断流程可以高频率地反复运行，从而为下游滚动时域运动规划实时生成更新后的后验分布。

5 实验

本节介绍用于检验以下关于所提出的贝叶斯逆博弈框架的假设的实证结果：

- **(H1，推断质量)**：与基于最大似然估计的逆博弈方法相比，所提出的框架能够更准确地推断未知目标函数，尤其是在多个目标函数都能解释同一组观测的情况下。
- **(H2，多模态不确定性感知)**：该框架能够量化目标推断中的不确定性，并捕捉学习到的先验分布和后验分布中的多模态结构。
- **(H3，规划安全性)**：在下游规划中使用推断得到的后验分布，与基于最大似然估计逆博弈解进行规划相比，能带来更安全的机器人行为。
- **(H4，多模态观测)**：该框架支持整合额外的观测模态（例如，将视觉线索与部分状态轨迹结合），以进行目标推断。
- **（H5，轨迹历史有限或缺乏信息）**：当轨迹历史有限（例如，智能体刚进入场景）或缺乏信息时，融合多模态观测有助于推断未知的对手意图，并降低后验分布的不确定性。
- **（H6，安全性与舒适性）**：多模态观测降低不确定性并更早推断意图，从而提高规划安全性，并通过避免不必要的转向提升运动舒适性。

5.1 实验设置与基线方法

在我们的主要评估实验中（第5.2节和第5.4节），我们考虑一个双智能体交叉路口场景（参见图1）在该场景中，自车机器人与一名对手进行交互，而自车并不知道对手的意图。除非另有说明，对手的真实意图来自一个权重相等的双分量高斯混合，其两个模式分别对应于在交叉路口左转或直行。

作为一项辅助研究，第 5.3 节对一个简化的高速公路场景进行评估，以更直观地定性展示所提出方法推断出的后验分布。高速公路场景的设置详见第 5.3 节。

为模拟策略性交互，我们使用智能体的真实意图，以滚动时域方式求解轨迹博弈，从而生成对手的行动。自车智能体则通过求解相同的滚动时域轨迹博弈进行控制，但使用的是由以下部分详述的评估方法推断出的对手意图第5.1.1节。

我们使用运动学自行车模型对每个智能体进行建模，其状态为 $x_t^i = (p_{x,t}^i, p_{y,t}^i, v_t^i, \xi_t^i)$，在时间步 t 的控制为 $u_t^i = (a_t^i, \eta_t^i)$，其中 $(p_{x,t}^i, p_{y,t}^i)$ 表示位置，v_t^i 表示纵向速度，ξ_t^i 表示航向角，a_t^i 表示加速度，η_t^i 表示转向角。在第5.2节中，智能体由分散式滚动时域博弈解直接控制。在第 5.4 节中，智能体由 CARLA 仿真器（Dosovitskiy 等，2017）中的低层 PID 控制器控制，以跟踪相应的分散式滚动时域博弈解。

为清楚起见，我们将自车智能体编为 $i = 1$，将对手编为 $i = 2$。在每个时间步，每个智能体都会在长度为 $T = 15$的规划时域内最小化其成本：

$$J_\theta^i = \sum_{t=1}^{T-1} \|p_{t+1}^i - p_{\text{goal}}^i\|_2^2 + 0.1\|u_t^i\|_2^2 + 400 \max\left(0,\ d_{\min} - \|p_{t+1}^i - p_{t+1}^{\neg i}\|_2\right)^3, \quad (14)$$

其中，$p_t^i = (p_{x,\ t}^i,\ p_{y,\ t}^i)$ 表示智能体 i 在时间t的位置，p_{goal}^i 是智能体 i 的目标位置，$d_{\min}$ 是允许的最小智能体间距离，θ 表示对手的潜在意图。该目标函数体现了到达目标、控制代价与避碰之间的权衡。在此示例中，未知的对手意图参数 θ 对应于对手的二维目标位置。

在 第5.2节中，自车机器人接收对手的部分状态观测，包括其过去 15 个时间步的位置和朝向。在 第 5 节中，我们还评估了所提方法的一种变体，该变体在推断时纳入对手的一帧原始图像。我们从原始图像中提取768维视觉特征，使用预训练的 DINOv3 ViT-B/16 模型（Sim´eoni 等人，2025），并将这些特征输入结构化变分自编码器。

结构化变分自编码器的编码器 e_ψ 和解码器 d_ϕ 均采用全连接前馈网络实现，包含两个隐藏层，如第 4 节所述。在第5.2节中，我们使用16维潜变量 z，编码器和解码器的隐藏层宽度分别为128 和80 。在第5.4节中，为适应更高维的观测，我们使用一个64维潜变量 z，编码器和轨迹解码器的隐藏层宽度分别为512和320，图像解码器的隐藏层宽度则为512 。

我们使用由闭环游戏过程生成的交互数据训练结构化变分自编码器，不使用参与者意图标签。具体而言，我们通过反复以滚动时域方式求解真实博弈来模拟交互，其中对手的真实目标 θ 从上文所述的双模态高斯混合中采样。对于每个交互回合，我们沿轨迹滑动一个长度为 15 个时间步的窗口，构造训练样本，形成包含智能体位置和朝向的观测序列。

对于 第5.4节中的多模态结构化变分自编码器，我们还将每段轨迹观测序列与窗口结束时捕获的俯视 RGB 图像帧配对，以反映对手最近的配置。在 第5.2节中，我们使用 560 个交互回合进行训练；在第5.4节中，我们使用 700个交互回合，以应对该设置日益增加的复杂性。我们使用 Adam（Kingma 和 Ba，2015）在一台 32 核台式机上训练每个结构化变分自编码器，实际运行时间约为 14 小时。

5.1.1 基线方法 我们在运动规划任务中评估贝叶斯逆博弈框架：自主机器人与意图未知的对手智能体进行交互。在每个步骤中，自车机器人都会接收对手的新观测，并推断对手的意图。获得该逆博弈解后，自车机器人通过求解由推断出的目标函数参数化的相应博弈，应用标准的博弈论运动规划方法，以计算其滚动时域行动。我们比较以下方法组合：

(i) 真实值（GT）。该方法作为一个“预言机”，能够访问对手智能体的真实意图。

(ii) 贝叶斯逆博弈（我们的方法） +期望规划（B-PinE）。该方法首先使用所提出的方法求解贝叶斯逆博弈，然后通过最小化期望代价 $\mathbb{E}_{\theta\sim q_{\psi,\ \phi}(\theta|y)}\left[J_\theta^1\right]$，在应变规划框架（Peters 等人，2024）中为机器人进行规划，其中期望是基于推断出的目标函数后验分布计算的。

(iii) 贝叶斯逆博弈（我们的方法） + 最大后验规划（B-MAP）。此方法首先使用所提出的方法求解贝叶斯逆博弈，然后

ˆ 提取最大后验（MAP）估计值 $\theta_{\mathrm{MAP}} \in \arg\max_\theta q_{\psi,\ \phi}(\theta \mid y)$，并求解相应的博弈 $\hat{\Gamma}(\theta_{\mathrm{MAP}})$。

(iv) 随机初始化的最大似然估计规划（R-MLE）。该方法使用Liu 等人的在线梯度下降方法(2023年)，求解最大似然估计逆博弈 $\hat{\theta}_{\mathrm{MLE}} \in \arg\max_\theta p(y \mid \theta)$，然后在与上述方法相同的博弈结构下求解ˆ博弈 $\Gamma(\theta_{\mathrm{MLE}})$。在线逆博弈优化的初始猜测值从覆盖候选真实目标的矩形区域中均匀采样得到。

(v) 贝叶斯先验初始化最大似然估计规划（BP-MLE）。该方法与 R-MLE 完全相同，不同之处在于，最大似然估计逆博弈优化的初始化方式是从我们方法学习到的贝叶斯先验分布 中采样对手意图，而不是从均匀分布中采样。

(vi) 静态贝叶斯先验规划（St-BP）。该方法从学习到的贝叶斯先验分布中采样对手意图，并直接求解由采样目标函数参数化的博弈。此方法作为消融实验，用于单独评估执行逆博弈目标推断对下游规划性能的影响。

对于使用所提出的变分自编码器进行逆博弈推断的方法，我们在每个时间步抽取 1000 个样本，以近似后验分布。得益于摊销推断，该采样过程可实时运行，耗时约 7 毫秒，避免了先前逆博弈方法所需的繁重在线优化（Liu 等人（2023）；Peters 等，2023b）。对于 B-PinE，我们将后验样本聚为两组，以匹配 Peters 等 (2024) 中的多假设博弈求解器。我们指出，对于更复杂的信念分布，已有现代计算方法可用于并行化计算（Li 等 2023a）。

5.2 仅部分状态观测下的 *IntersectionScenario*

我们首先在一个仅有轨迹观测的模拟交叉路口驾驶场景中评估所提出的方法。图4 展示了我们的 B-PinE 方法（上）和 R-MLE 基线方法（下）所产生的代表性行为快照。

对于 B-PinE，在对手进入交叉路口之前，观测到的状态历史不足以唯一确定对手的意图——它是直行还是左转。因此，我们的方法推断出一个双峰后验分布，其中两种意图的可能性大致相等。随后，博弈论规划器据此对对手未来运动进行多峰预测，并采取保守计划以应对这两种可能：机器人让行，并从对手左侧驶过。随着对手接近交叉路口，其行为提供的信息越来越多，后验分布收敛为单峰分布，意图得以完全判别，机器人也成功解决了这次交互。

相比之下，在对手进入交叉路口之前，R-MLE 基线返回一个点估计，轻率地断定对手打算直行——这能完美解释观察到的行为，却忽略了先验知识：另一种意图也可能是合理的。由于缺乏对不确定性的认知，机器人（错误地）

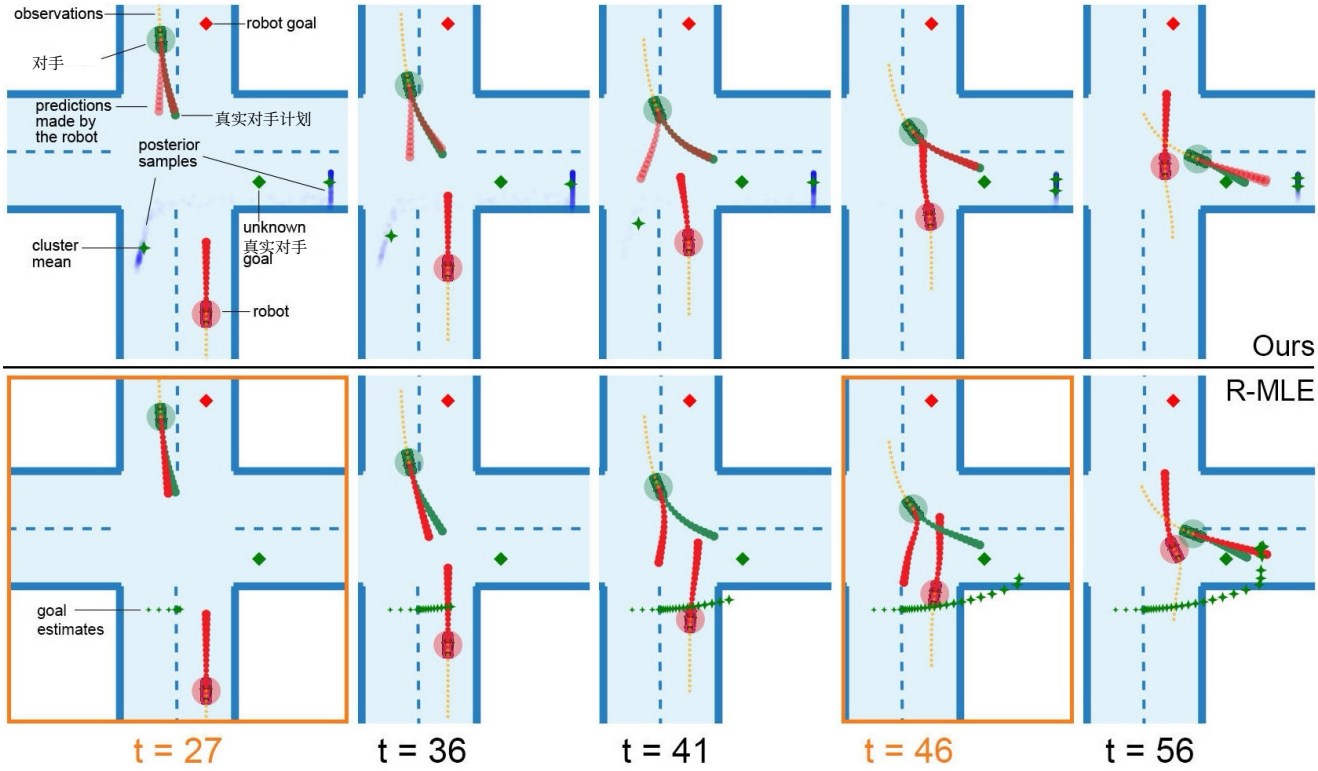


图4。 交叉路口场景的快照（第5.2节），其中自车机器人分别由我们的 B-PinE 方法（上）和 R-MLE 基线（下）进行控制。对于 R-MLE，估计目标显示为绿色星形，其大小会随时间增大。转载自会议论文（Liu 等，2024）。

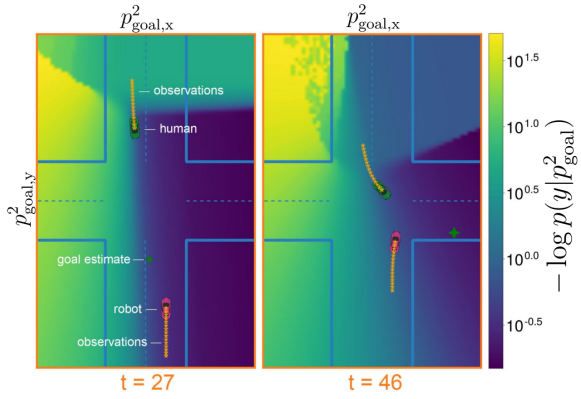


图5。 MLE逆博弈代价分布（负观测对数似然， $-\log p(y \mid p^2_{\text{goal}})$，关于对手目标位置 p^2_{goal}），展示了 R-MLE 基线在图图4所示交互过程的第27和第46个时间步。由于忽略了贝叶斯先验，MLE逆博弈问题可能是不适定的，并呈现平坦的代价分布；因此，其解可能导致不安全的后续决策（参见图4）。转载自会议论文（Liu 等，2024）。

对直行假设深信不疑，并计划激进地穿越交叉路口。最终揭示对手的真实意图时，最大似然估计的解转向了正确的目标函数；然而，机器人此时已经采取激进机动，无法及时调整以避免碰撞。

5.2.1 MLE 逆博弈的可观测性挑战为了更好地理解 R-MLE 的故障模式基线方法（见图4）图5 深入展示了 R-MLE 在线求解的底层 MLE 逆博弈问题，并将得到的 MLE 代价分布可视化，即观测对数似然的负值 $-\log p(y \mid p^2_{\text{goal}})$ 在对手目标位置 p^2_{goal}上的取值，展示了两个具有代表性的时间步。回顾一下，MLE 公式通过忽略先验分布并贪婪地最大化观测似然，简化了贝叶斯推断：

$$p(p^2_{\text{goal}} \mid y) \propto \underbrace{p(y \mid p^2_{\text{goal}})\, p(y)}_{\text{Bayesian}} \cancel{\propto} \underbrace{p(y \mid p^2_{\text{goal}})}_{\text{MLE}} \cancel{p(y)}.$$

因此，在对手进入交叉路口且其运动开始透露其意图之前，MLE 的代价分布可能近乎平坦（参见图5中的暗色区域）。在这种情况下，许多目标假设都能同样充分地解释观测，因此 MLE 求解器可能会返回一个实质上任意的点估计，从而导致不安全的后续决策。相比之下，我们的贝叶斯逆博弈方法明确纳入先验分布并量化后验不确定性，使机器人能够在对手意图不明确时采取保守的计划。

总体而言，这些定性行为支持假设 *H1* 以及 *H2* 中聚焦于后验分布的部分。

5.2.2 蒙特卡洛评估 我们通过蒙特卡洛研究对 第5.1.1节 中的方法进行定量评估。我们针对图4所示的交叉路口场景进行了 1500 次仿真试验。每次试验中，自车机器人的初始位置沿其车道均匀采样，从而生成多样化的初始条件和交互情形——从自车机器人先进入交叉路口的情况，到

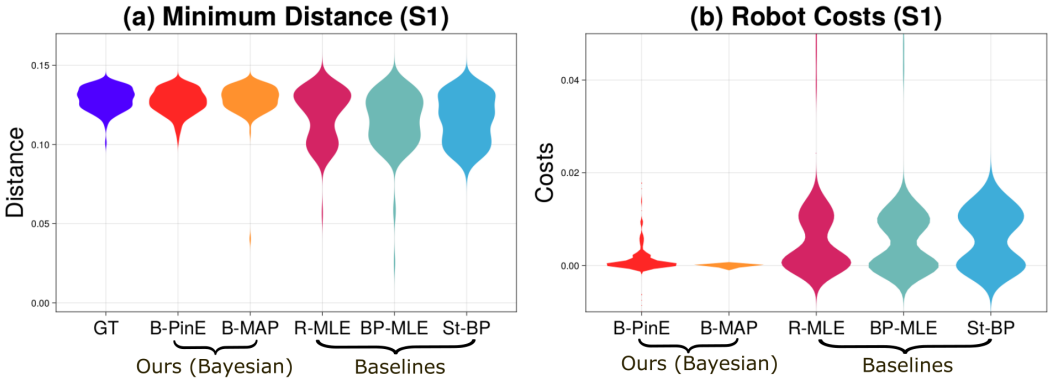


图6。 图4所示交叉路口场景（第5.2节）中自车机器人在左转对手车辆之后通过时的蒙特卡洛研究结果。与基于最大似然估计的逆博弈方法以及不求解逆博弈的基线方法相比，贝叶斯逆博弈方法在安全性（更高的最小智能体间距）和效率（更低的自车机器人代价）方面显著更优。机器人代价是在减去真实博弈代价后报告的。改编自会议论文（Liu 等，2024）。

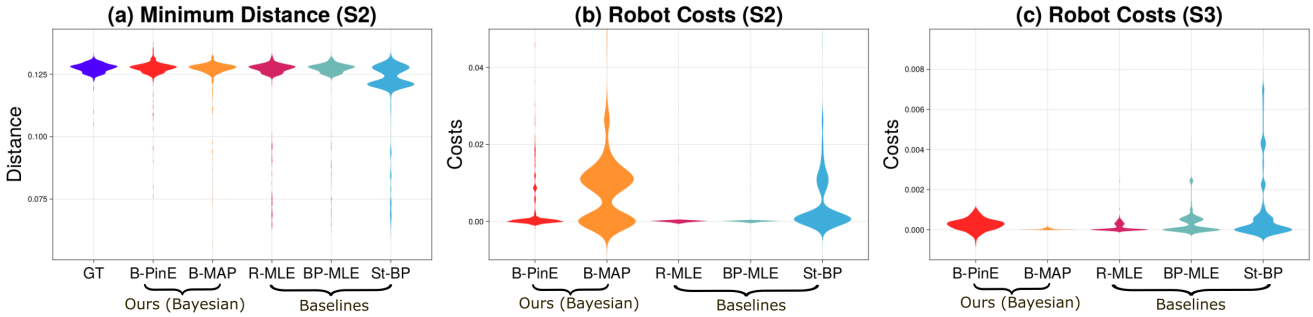


图7。 图4所示交叉路口场景（第5.2节）的蒙特卡洛研究结果：（a–b）对应自车机器人在左转对手车辆之前进入交叉路口的试验，（c）对应对手直行通过交叉路口且不与自车机器人互动的试验。改编自会议论文（Liu 等，2024）。

对手先到达的情况。我们报告两项指标：（i）整个仿真过程中的最小智能体间距，作为规划安全性的指标；（ii）自车机器人产生的代价，作为效率的衡量标准。

我们根据对手意图和初始条件，将1500次试验分为三类：**（S1）**对手左转并在自车机器人之前到达交叉路口。该安全关键场景中存在激进对手车辆；自车机器人必须及时推断对手意图并让行，以确保安全。**（S2）**对手左转，但自车机器人先到达交叉路口，可以从容通过。在此情况下，由于自车机器人具有时机优势，准确识别意图并不那么关键。**（S3）**对手直行通过交叉路口。满足安全性较为容易，但自车机器人理想情况下应尽早推断出这一意图，以避免过度保守的行为损害效率。

图 6 展示了 **(S1)** 中各项试验的蒙特卡洛结果。采用我们提出的贝叶斯逆博弈解的方法，其规划安全性显著更高（最小智能体间距更大），效率也更高（自车机器人代价更低）。在此场景中，求解最大似然估计逆博弈的表现大致与 St-BP 相当，也就是说，直接使用我们的贝叶斯先验进行规划，而不求解任何逆博弈，效果与之相近。这一结果凸显了准确掌握先验分布的重要性。以所有真实值试验中的最小智能体间距作为碰撞阈值，碰撞率分别为 **B-PinE 为 0.0%**，**B-MAP 为 0.78%**，R-MLE 为 17.05%，BP-MLE 为 16.28%，而 St-BP 为 17.83%。

St-BP。在 B-PinE 和 B-MAP 之间，B-PinE 使用完整后验分布进行规划，因此更加保守（安全性更高），但代价是机器人代价略高。

图7 （a–b）展示了 **（S2）**中各项试验的蒙特卡洛结果。正如预期，由于自车智能体先进入交叉路口，在这种情况下，我们的贝叶斯方法与基于最大似然估计的基线方法在安全性表现上的差异并不那么明显。尽管如此，我们的贝叶斯方法仍能提高安全性：使用与 **（S1）**相同的碰撞距离阈值时，碰撞率为**B-PinE 为 0.86%**，**B-MAP 为 2.24%**，R-MLE 为 7.59%，BP-MLE 为 6.03%，St-BP 为 7.59%。

在规划效率方面，B-MAP 仅依赖完整贝叶斯后验分布的点估计，因此会采取过度自信且激进的机动，导致智能体之间发生协调故障（例如，双方都试图同时进入交叉路口，随后不得不刹车）。相比之下，我们方法的 B-PinE 变体利用完整后验分布进行考虑不确定性的规划，从而实现更高的效率（和安全性），凸显了基于完整后验分布进行规划的重要性。

最后，图7（c）显示，在 **（S3）**中，B-MAP 实现了最高的规划效率，而 B-PinE 按其构造更加谨慎，与基于点估计的方法相比，会产生少量略为保守的试验。在这种情况下，非逆向博弈基线方法 St-BP 的表现明显较差。

综合来看，这些结果表明我们的贝叶斯逆博弈方法能够提升下游

运动规划的安全性优于基线方法，并支持假设 *H3*。在两种贝叶斯变体 B-PinE 和 B-MAP 中，我们观察到，使用完整贝叶斯后验分布时，（**S1–S2**）的规划安全性有所提升，而（**S2**）的效率有所提高。

5.3 仅部分状态观测下的高速公路场景

第5.2节评估了所提出的贝叶斯逆博弈框架与下游运动规划方法相结合的效果，检验求解完整贝叶斯问题对整体规划流程的益处。本节随后考虑一个简化场景，并进一步聚焦于单独考察我们的方法所生成推断分布的定性行为，不涉及规划任务。

我们考虑一个高速公路驾驶场景，其中两个智能体依次行驶在单车道道路上。每个智能体都有一个期望速度，但前方智能体可以自由地按照其目标速度行驶，而后方智能体则负责避免碰撞——如果它想比前方智能体开得更快，就必须减速，因此其真实意图速度无法观测。

通过使用两个智能体的真实目标速度求解博弈，对它们的行为进行联合仿真。我们的方法观测两个智能体的轨迹，并在假设已知前方智能体期望速度的前提下，推断后方智能体的期望速度。根据构造，在后方智能体被前方智能体阻挡、其真实期望速度无法观测的情况下，我们预期所提出的贝叶斯方法会得到不确定性较高的后验分布。

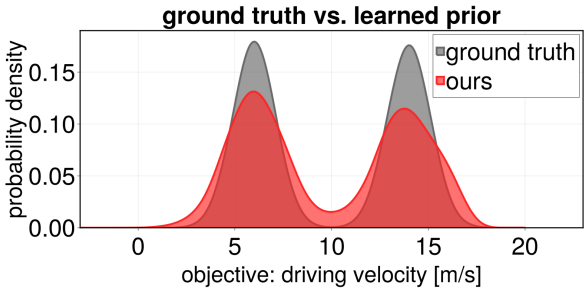
在这个简化的高速公路场景中，每个智能体都被建模为一个双积分器：纵向位置和速度构成状态，加速度则是控制输入。为推断对手的潜在意图，我们使用一个变分自编码器，其编码器/解码器隐藏层大小与第5.2节中的相同，但采用一维潜变量。观测由两个智能体连续 15 步的速度组成。

我们使用包含 20,000 条模拟观测的数据集训练结构化变分自编码器模型。在每次试验中，自车智能体的目标速度从0 米/秒$^{-1}$ 到20 米/秒$^{-1}$之间均匀采样，而对手的期望速度则从双峰高斯混合分布中采样（图 8上方的灰色部分），该分布由两个方差为 1 的分量构成，其均值分别位于最大速度的 30% 和 70% 处。

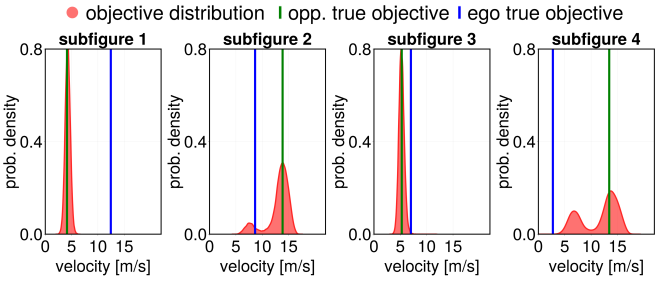
图 8a 展示了从智能体轨迹数据集中学习到的先验分布，该数据集未标注后方智能体的意图。通过解码先验潜在空间中的样本即可得到该先验分布。所提出的贝叶斯框架成功恢复了与真实值高度吻合的双峰先验分布；先前的实验已强调其重要性，参见图5。

图 8b 展示了推断得到的后验分布的代表性快照。当前方智能体的期望速度高于后方智能体时，我们的方法会产生一个尖锐的单峰后验分布，能够以较低的不确定性正确捕捉后方智能体的真实意图。相比之下，当后方智能体被前方智能体阻挡时——也就是说，后方智能体的期望速度更高，但受到前方智能体的限制——我们的方法会产生一个宽广的双峰后验分布，且接近先验分布，表明不确定性较高。

结合第 5.2 节中的结果，这些发现支持假设 *H2*：我们的贝叶斯逆博弈框架具有不确定性感知能力，并且能够捕捉先验和后验分布中的多模态性。



通过解码先验潜变量样本获得的先验分布 .



通过解码后验潜变量样本获得的后验分布。

图 8。 在高速公路驾驶场景中由我们的贝叶斯逆博弈方法生成的先验和后验分布（第 5.3 节）。该方法成功地从无标签数据中学习先验，并在后验分布中体现出不确定性感知能力：当观测能够识别隐参数时，模型会产生尖锐的分布；而当观测缺乏信息时，则会产生接近先验分布的宽广多模态分布。改编自会议论文（Liu 等，2024）。

5.4 IntersectionScenario 与多模态观测

本节根据第 4.1.1 节中的讨论，将我们的贝叶斯逆博弈框架的评估扩展到每个样本包含多个观测模态的情形，从联合图像和部分状态轨迹观测中推断未知博弈参数。

为了评估该方法在照片级真实感视觉输入下的效果，我们在 CARLA 仿真器（Dosovitskiy 等，2017）中研究了两个变体——图像-轨迹变分自编码器和仅轨迹变分自编码器。在这些实验中，车辆由 PID 控制器驱动，以跟踪分散式滚动时域博弈解。

我们在图4的交叉路口场景中评估两个示例，使用的数据集中的图像观测通过对手的颜色（第5.4.1节）或车辆类型（第 5.4.2 节）提供关于对手意图的不同程度的上下文信息。在第5.4.1节中，图像信息量很大，能够有效确定对手隐藏目标的模态（直行或左转）；而在第 5.4.2 节中，图像仅提供

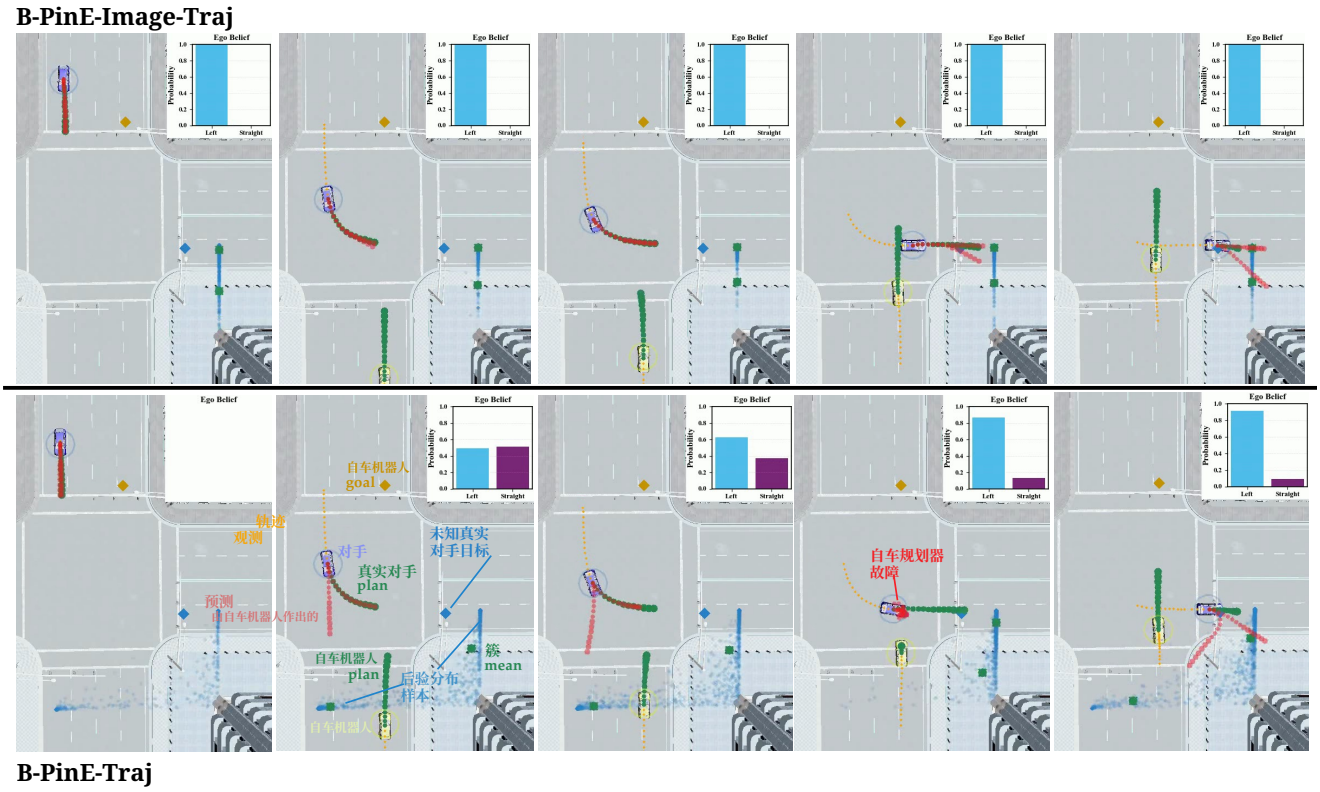


图9。 交叉路口场景（第5.4.1节）中的快照，场景位于 CARLA仿真器中，对手意图通过其颜色编码。由 B-PinE 与图像-轨迹变分自编码器控制的自车智能体（上图）利用视觉上下文信息，及早识别出对手的左转意图，并安全地礼让以解决交互问题。相比之下，使用仅轨迹变分自编码器的 B-PinE（下图）在图4中展示，过晚才识别出对手意图，导致规划器故障，并触发不安全的紧急急刹车。

部分信息，对轨迹观测起到有限的补充作用。

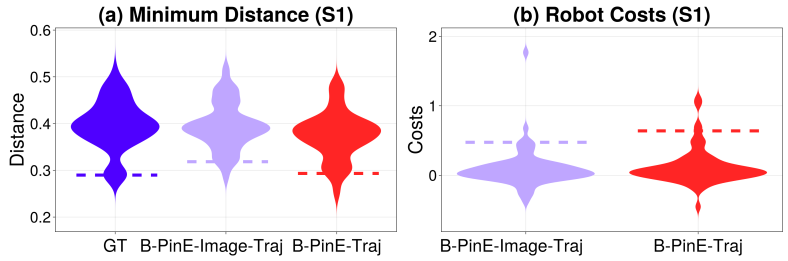
5.4.1 颜色编码的智能体意图 本节评估所提出的框架，设置中对手的颜色提供有关其未知意图的信息。我们构建了一个数据集，其中，对手有50%的概率是向左转的蓝色汽车，有50%的概率是直行的红色汽车，如图9所示。这一设置模拟了现实场景中的情况：对手会明确示意其意图，例如通过转向灯示意。随后，我们训练了两个结构化变分自编码器实例：一个同时使用对手图像和其过往行为的部分状态观测（如上文所述），另一个仅使用部分状态观测。

在图像-轨迹变分自编码器中，图像观测为后验推断提供了额外的上下文。当图像包含轨迹所能提供的信息之外的内容时（如本例所示），基于两种模态推断出的后验分布应当比仅基于轨迹数据推断出的后验分布具有更低的不确定性，并且更接近对手的真实意图。

定性行为。我们使用与第5.2节中相同的 B-PinE 规划器，并比较其与上述两种结构化变分自编码器之一结合时的闭环性能。图9展示了这两种变体可能产生的定性行为差异。当对手首次进入场景、尚无轨迹历史可用时，图像-轨迹变分自编码器能够成功仅凭图像便利用了丰富的先验信息（对手颜色），并结合训练期间学到的轨迹统计信息，计算出关于对手未知目标的范围较窄且准确的后验分布（第1帧）。相比之下，此时仅轨迹变分自编码器缺乏足够的信息，无法形成后验分布。随着交互展开、更多轨迹数据逐渐可用，图像-轨迹变分自编码器仍能正确推断对手正在左转，并让自车做好让行准备。然而，仅轨迹变分自编码器最初产生了双峰后验，对手是直行还是左转仍不确定；直到后来才收敛到左转模式，而此时 B-PinE 规划器已经决定从右侧积极超车（第2至第4帧）。因此，搭配图像-轨迹变分自编码器的 B-PinE 规划器能够安全让行并顺利化解交互；而搭配仅轨迹变分自编码器的 B-PinE 规划器则会过于接近迎面驶来的车辆，博弈论规划器也无法找到可行解；最终，自车采取了非预期的急刹车。

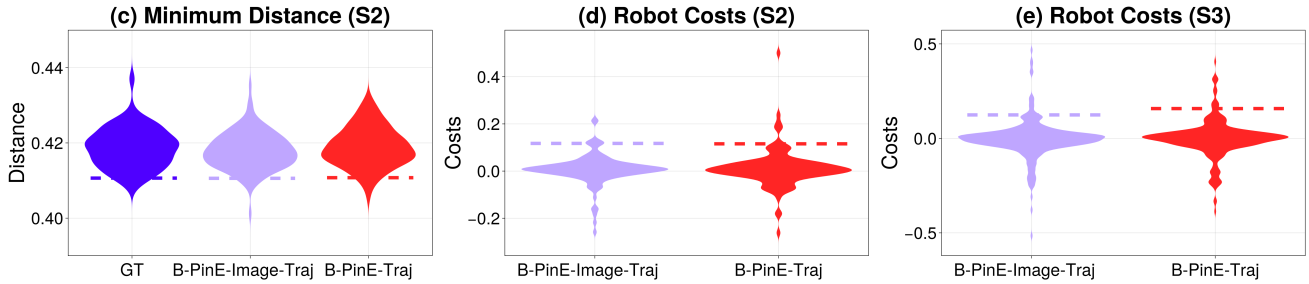
这一交互支持我们的假设*H4-5*，并突出了我们的贝叶斯逆博弈框架在整合多模态观测进行推断时的两个关键优势：

- 它可以利用视觉模态中的上下文先验信息，立即形成后验分布
- 当仅凭轨迹信息不足以判断意图、否则会导致高度不确定性时，它可以消除歧义。



自车机器人在左转对手车辆通过交叉路口后驶过交叉路口的试验

。



对手车辆左转且自车先行（c–d），或对手车辆直行（e）的试验。

图10。 图9所示交叉路口场景的 蒙特卡洛 (第5.4.1节)结果。(a、c)图展示试验 **(S1–2)**的最小智能体间距，(b、d、e)图展示 **(S1-3)** 的自车机器人代价。代价以真实成本为基准表示。虚线分别表示最小距离的第5百分位数和代价的第95百分位数。在贝叶斯逆博弈中联合使用图像和轨迹观测（B-PinE-Image-Traj），可进一步提升 **(S1)**中的规划安全性（最小智能体间距更大，碰撞率更低——B-PinE-Image-Traj为0.0%，而B-PinE-Traj为1.57%），且不牺牲规划效率（代价相当）。此外，B-PinE-Image-Traj通过避免不必要的转向力度，提升了运动舒适性；参见 图11。

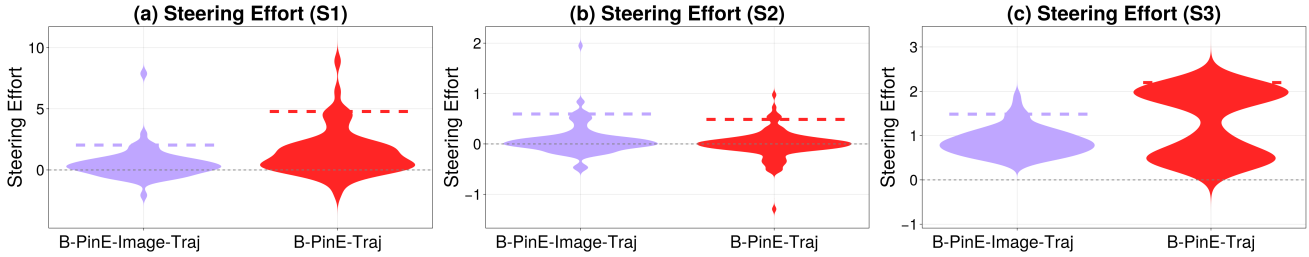


图11. 求解基于图像—轨迹观测的贝叶斯逆博弈与仅基于轨迹观测的贝叶斯逆博弈时的转向力度，并以求解真实博弈为参照。通过整合多模态观测，图像-轨迹变分自编码器能够及早消除对手意图的歧义，避免不必要的转向——在（**S1**）中效果适中，在（**S3**）中效果显著。虚线表示第95百分位数。

这些特性共同使下游决策更加安全，适用于瞬息万变、瞬间演化的多智能体交互。

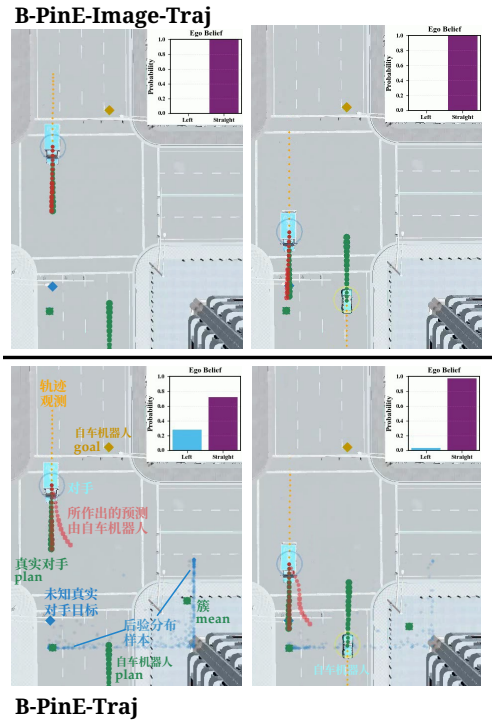
蒙特卡洛评估。 接下来，我们通过一项蒙特卡洛研究，对贝叶斯逆博弈框架的两个变体进行定量评估。具体而言，我们比较了与 B-PinE 下游规划器结合使用时的两个已训练变分自编码器。每种方法均运行 500 次仿真，初始条件和对手意图的随机化方式与上述训练数据中的方式相同。

总的来说，我们从这项蒙特卡洛研究中得出两项主要观察结果：

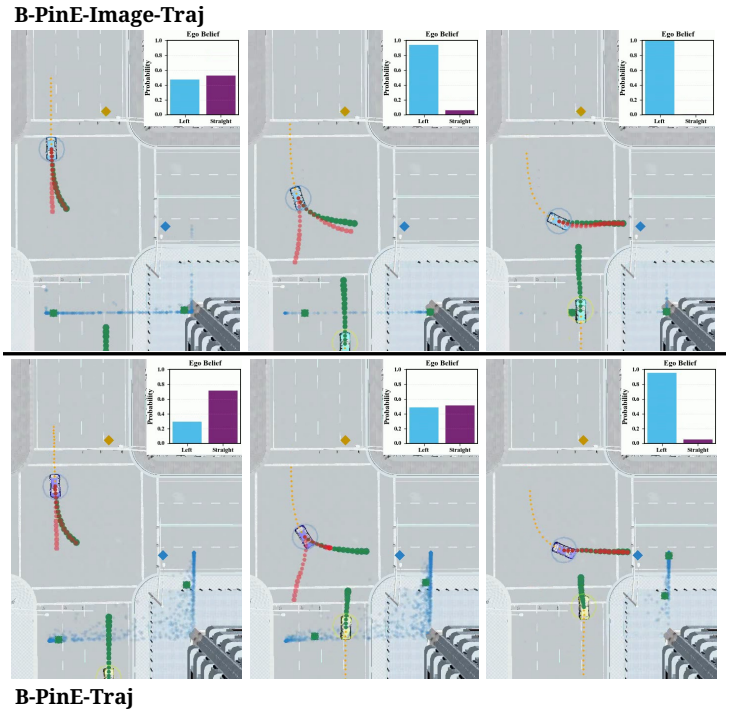
- **规划效率与安全性**。与仅轨迹变分自编码器相比，使用我们的图像-轨迹变分自编码器进行规划，能够进一步提升下游规划的安全性，同时不会牺牲规划效率，这一点可从博弈代价中体现出来。总体而言，在大多数运行中，使用我们的贝叶斯逆博弈方法的任一变分自编码器变体进行规划，所得成本表现都相近，并且两者都接近假设已知智能体意图时的真实值博弈成本。
- **转向力度**。 在对手直行的情况下，图像-轨迹变分自编码器能够及早消除对手意图的歧义，从而减少不必要的转向力度并提升运动舒适性。

我们按照与第5.2节相同的方式对蒙特卡洛试验进行分组。图10报告了 **(S1)** (a–b)、**(S2)** (c–d) 和 **(S3)**(e) 的成本和最小智能体间距结果。

总体而言，图像-轨迹变分自编码器与轨迹变分自编码器在效率（成本）和安全性（最小智能体间距）方面表现相当。不过，在 **(S1)**(图10a) 中，图像-轨迹变分自编码器通常比仅轨迹变分自编码器实现更大的最小智能体间距（参见图像-轨迹变分自编码器较高的第5百分位数，



对手智能体是一辆卡车。



对手智能体是一辆汽车。

图12。 交叉路口场景中的快照（第 5.4.2 节），其中对手可能是卡车，也可能是汽车；由于高度限制，卡车禁止左转。当对手是卡车时，图像-轨迹变分自编码器利用视觉输入中的车辆类型线索，立即推断出对手的意图（直行）；而仅轨迹变分自编码器则要等到之后根据观测到的运动（对手驶过交叉路口后）才能识别这一意图。当对手是汽车时，两个变分自编码器的表现相似，最初都会生成双峰后验分布，并随着交互展开而收敛为单峰后验分布。

以虚线表示），这表明在不牺牲规划效率的情况下提升了安全性（规划成本相当也反映了这一点）。安全裕度的提升直接转化为更低的碰撞率：在蒙特卡洛研究中，采用与第5.2节相同的碰撞距离阈值，**(S1)** 的碰撞率分别为 B-PinE-Image-Trajectory 的 0.0% 和 B-PinE-Trajectory 的 1.57%；而在 **(S2)** 中，B-PinE-Image-Trajectory 和 B-PinE-Trajectory 的碰撞率均为 0.0%。需要指出的是，**(S1)** 对安全性要求更高：自车在激进对手车辆之后驶入交叉路口，并且必须判断对手是否有意让行。

在 图10所示的博弈代价中，与任务相关的项（例如，到目标点的距离和接近惩罚）严格占主导地位，控制 effort 惩罚则相形见绌；参见 公式（14）。然而，对于自动驾驶而言，从乘客的角度看，运动舒适性同样重要，甚至更加重要。 图11 显示，配备图像-轨迹变分自编码器的 B-PinE 比配备仅轨迹变分自编码器的 B-PinE 更能有效避免不必要的转向操作——在 **(S1)** （图11a）中效果适中，在 **(S3)** （图11c）中效果显著；在后一种情况下，对手直行且不与自车交互。这是合乎直觉的：对于仅轨迹变分自编码器，在其对对手意图的不确定性降低之前，它往往会采取保守策略，同时防范两种可能性（对手直行或左转）。相比之下，图像-轨迹变分自编码器能及早正确识别对手意图，并从容直行，避免不必要的转向。

我们强调，此处使用仅轨迹变分自编码器的 B-PinE 较为保守，并不与其在图9中的激进表现相矛盾。正是因为图像-轨迹变分自编码器能够较早推断出对手意图，所以当对手实际向左转时，它可以采取保守策略；与此同时，当对手也直行时，它又能轻松地直行。

综合来看，这些结果支持我们的假设*H6*。

5.4.2 由智能体类型编码的意图 第5.4.1节 研究了一个示例，其中图像观测对于推断对手智能体的意图极具信息量，例如根据转向信号推断驾驶员意图的场景。在这里，我们研究一种情形：图像包含的直接信息较少，但仍有助于推断对手意图。

我们评估与第5.4.1节相同的交叉路口场景，但随机改变对手颜色，使其无法提供有关意图的信息。相反，可以根据对手的车辆类型部分推断其意图：对手要么是汽车，要么是卡车。在这个示例中，由于高度限制，卡车不允许左转。我们构建训练数据集，使对手有80% 的概率为汽车，有20% 的概率为卡车。对手是卡车时，只能直行；对手是汽车时，左转或直行的概率各为50%。除这些变化外，我们采用与第5.4.1节相同的变分自编码器训练设置。

在这种情况下，训练良好的图像-轨迹变分自编码器应能立即推断出卡车对手将直行；而当对手是汽车时，其推断结果在定性上应与仅轨迹变分自编码器相似。

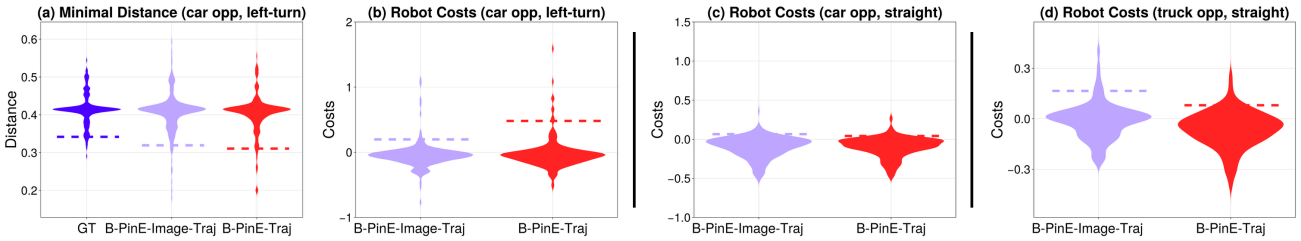


图13。图12中交叉路口场景的（第 5.4.2 节）蒙特卡洛结果。虚线表示最小距离的第5百分位数和代价的第95百分位数。面板 (a,b) 展示对手为左转汽车的情况；面板 (c) 展示对手为直行汽车的情况；面板 (d) 展示对手为直行卡车的情况。代价以真实成本为基准显示。在各场景中，我们方法的两个变体取得了相近的效率表现（代价）；而在 (a) 中，纳入图像—轨迹观测提高了安全性（碰撞率：B-PinE-Image-Traj 为 0.58%，B-PinE-Traj 为 1.73%）。此外，当对手是被限制为直行的卡车时，B-PinE-Image-Traj 通过避免不必要的转向改善了运动舒适性；参见图14。

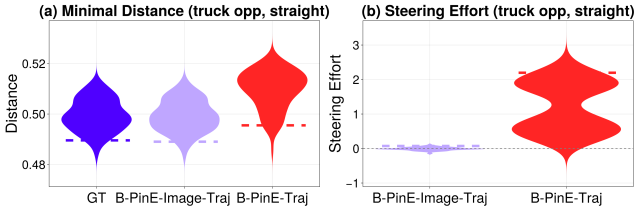


图14。当对手为直行卡车时，我们的贝叶斯逆博弈方法的两个变体的最小距离和转向力度（第 5.4.2 节）。转向力度以求解真实博弈时的结果为基准显示。虚线表示最小距离的第5百分位数和转向力度的第95百分位数。通过根据对手类型的图像观测，及早推断对手的意图（直行），图像-轨迹变分自编码器通过避免不必要的转向改善了运动舒适性。相比之下，仅轨迹变分自编码器更加谨慎——在迎面而来的对手驶过交叉路口之前始终转向远离对手（从而产生更大的最小距离），因此需要更大的转向力度。

定性行为。如图12所示，当对手车辆是卡车（左图）时——由于高度限制，卡车不能左转——图像-轨迹变分自编码器会利用视觉输入中的车辆类型线索，立即推断其意图（直行）。这是因为在我们的交互数据集中，卡车在这个交叉路口从不左转，而这一限制只能通过图像模态观察到。相比之下，仅轨迹变分自编码器最初会产生双峰后验，只有在对手几乎通过交叉路口后，该后验才会收敛为（近乎）单峰。

当对手车辆是汽车时（图12，右图），两个变分自编码器的表现相似：二者在早期都推断出双峰后验分布，随着交互展开，该后验分布逐渐收缩为单峰后验分布，反映出对对手意图的确定性不断提高。

仅轨迹变分自编码器产生了不平衡的初始后验分布，因为它无法观测车辆类型；由于卡车只能直行，在这种情况下直行的总体概率更高。

这些结果支持我们的假设 *H4-5*，并凸显了贝叶斯逆博弈框架的一项实用特性：它能够整合不同观测模态的信息。当图像提供额外的有效线索时，该框架会利用这些线索补充基于轨迹的推断；当图像提供的信息有限时，它则会自然地主要依赖轨迹观测。

蒙特卡洛评估。接下来，我们通过一项蒙特卡洛研究，对贝叶斯逆博弈框架的两种变体进行定量评估。与第 5.4.1节一样，我们比较这两种经过训练的变分自编码器与 B-PinE （作为下游规划器）结合时的表现。我们对每种方法进行 500 次仿真，并按照上述方式随机设置初始条件、对手类型和对手意图。

我们将试验分为三种情景：(i) 对手是左转汽车，(ii) 对手是直行汽车，(iii) 对手是直行卡车。在 (i) 和 (ii) 中，我们预计两种变分自编码器变体的表现相近，因为观察到对手是汽车，并不能让图像-轨迹变分自编码器明确判断其意图。相比之下，在 (iii) 中，图像-轨迹变分自编码器可以较早推断出对手意图——因为卡车受高度限制，只能直行——因此其行为应当不那么保守（例如，所需转向较少）。

图13报告了最小距离和自车机器人代价。总体而言，按机器人代价衡量时，两种变分自编码器变体的效率相当；而与仅轨迹变分自编码器相比，图像-轨迹变分自编码器提升了安全性：在对手汽车左转的试验中，B-PinE-Image-Traj 的碰撞率为 0.58%，B-PinE-Traj 为 1.73%。

此外，图14聚焦于对手是直行卡车的情况，并报告了两种变体的最小智能体间距和转向力度。与图12中的定性结果一致，图像-轨迹变分自编码器能够较早推断对手意图，行为也不那么保守，因此最小距离更小、转向力度更低。相比之下，轨迹变分自编码器通常要等卡车通过交叉路口后才能识别其意图（参见图12）；因此，尽管没有发生交互，B-PinE-Traj 仍会转向避开迎面而来的对手，导致最小距离更大、转向力度更高。总体而言，在这种情况下，通过多模态观测及早识别意图能够提升运动舒适性。

定量结果与第 5.4.1 节中的结果相似，支持我们的假设 *H6*。

6 结论

我们研究了不完整博弈模型下用于交互式决策的逆博弈。我们引入了一种结构化变分自编码器框架，该框架能够近似求解贝叶斯逆博弈，并根据包括高维图像在内的多模态观测，推断未知博弈参数（例如对手目标函数）的完整后验分布。该框架在变分自编码器中嵌入了可微纳什博弈求解器，提供了一种保留可解释性的归纳偏置，并使后验样本立足于博弈参数空间。

我们在交互式决策场景中对所提出的方法进行了广泛评估。与只能给出博弈参数点估计的常见基于最大似然估计的逆博弈方法相比，我们的框架能够有效量化不确定性、充分利用离线数据集中的先验信息，并在交互式运动规划中实现更安全的下游决策。在没有目标函数标签的交互数据上离线训练结构化变分自编码器后，该模型支持在测试时实时进行后验采样，无需额外求解博弈，因此适用于在线决策。最后，通过在不同模态间学习共享的潜在空间，所提出的方法能够在轨迹历史有限时利用上下文视觉信息，例如，当新的智能体进入场景且几乎没有或完全没有其过往行为数据时。

我们的当前实现使用一个解码器，将潜变量 z 确定性地映射到博弈参数 θ，这与标准变分自编码器的形式一致（Kingma 和 Welling 2014）。一个自然的扩展方向是引入随机解码机制，例如通过分层潜变量模型或基于扩散的生成模型，以更充分地发挥现代生成架构的表达能力。其次，我们流程的模块化结构便于研究其他均衡概念，例如熵成本均衡（Mehr 等，2023），这类概念可以捕捉有限理性假设。

作者贡献

Yash Jain: 方法学、数据整理、调查、软件、初稿撰写、审阅与编辑

XinjieLiu: 概念化、数据整理、形式分析、调查、方法学、软件、初稿撰写、审阅与编辑

Lasse Peters: 概念化、形式分析、方法学、初稿撰写、审阅与编辑

DavidFridovich-Keil: 经费获取、概念化、监督、初稿撰写、审阅与编辑

UfukTopcu: 经费获取、概念构思、监督、撰写——初稿、撰写——审阅与编辑

声明与声明事项

伦理考量

本文不包含任何涉及人类或动物参与者的研究；所有实验均在仿真环境中进行。

参与同意不适用。

发表同意

不适用。

利益冲突声明

作者声明，在本文的研究、撰写和/或发表方面不存在任何潜在利益冲突。

经费声明

本研究由美国陆军研究实验室通过合作协议 W911NF-23-2- 0011、W911NF-25-2-0021 和 W911NF-23-S-0001 资助，并由美国国家科学基金会通过拨款 2211432、2211548 和 2336840 资助。

数据可用性

可微博弈求解器的实现已公开发布（JGTP 团队，2023年），复现结果所需的实验代码也将通过同一 GitHub 组织发布。

AI 工具的使用

在准备本稿件期间，作者使用 ChatGPT（OpenAI）和 Gemini （ Google ）作为辅助工具，以提升语言表达的清晰度、语法和整体写作质量，并协助审查和调试研究软件（例如，识别潜在的代码错误并提出改进建议）。作者对所有 AI 辅助生成的内容进行了严格审阅和编辑，并在必要时核实其技术正确性及引用材料；作者对稿件及其参考文献的准确性、有效性和原创性承担全部责任。作者还承认大型语言模型存在局限性，包括可能产生偏见、错误或遗漏，并确认所有 AI 生成内容均经过适当的人工核验后才纳入稿件。

参考文献

Armstrong C、Park R、Liu X、Gupta K 和 Fridovich-Keil D（2025）推断动态非合作博弈中的前瞻性。*IEEE*机器人与自动化快报10(12)：13193–13200。DOI:10.1109/LRA.2025.3625521。Awasthi C 和 Lamperski A（2020）具有混合不等式约束的逆微分博弈。载于：*IEEE*美国控制会议（*ACC*）论文集。Bas,ar T 和 Olsder GJ（1999）动态非合作博弈论。第2版。工业与应用数学学会（SIAM）。

Billups SC、Dirkse SP 和 Ferris MC（1997）大规模混合互补问题求解器的比较。计算优化与应用7(1)：3–25。Clarke S、Dragotto G、Fisac JF 和 Stellato B（2023）学习势博弈中的理性。载于：*2023*年第*62*届*IEEE*决策与控制会议（*CDC*）。IEEE，第4261–4266页。Cleac'h L、Schwager M 和 Manchester Z（2022）ALGAMES：用于约束动态博弈的快速增广拉格朗日求解器。自主机器人46(1)：201–215。Diehl C、Klosek T、Krueger M、Murzyn N、Osterburg T 和 Bertram T（2023）用于联合运动预测与控制的基于能量的势博弈。载于：机器人学习会议（*CoRL*）论文集。Dirkse SP 和 Ferris MC（1995）PATH 求解器：一种用于混合互补问题的非单调稳定化方案。优化方法与软件5(2)：123–156。Dosovitskiy A、Ros G、Codevilla F、Lopez A 和 Koltun V（2017）CARLA：一个开放的城市驾驶仿真器。载于：第一届机器人学习年度会议论文集。第1–16页。Facchinei F 和 Pang JS（2003）有限维变分不等式与互补问题。施普普林格出版社。
Fridovich-Keil D、Ratner E、Peters L、Dragan AD 和 Tomlin CJ（2020）用于非线性多参与者一般和微分博弈的高效迭代线性二次近似。载于：*IEEE*国际机器人与自动化会议（*ICRA*）论文集。Geiger P 和 Straehle CN（2021）利用隐式层学习多智能体轨迹的博弈论模型。载于：人工智能进展会议（*AAAI*）论文集，第35卷。Hu H 和 Fisac JF（2022）人机交互中的主动不确定性降低：一种隐式双重控制方法。载于：机器人学算法基础国际研讨会（*WAFR*）。施普林格，第385–401页。Hu H、Fisac JF、Leonard NE、Gopinath D、DeCastro J 和 Rosman G（2025）深思且快速行动：从逆动态博弈中学习神经非线性意见动力学，以实现瞬时交互。载于：*2025 IEEE*国际机器人与自动化会议（*ICRA*）。IEEE，第16678–16684页。Inga J、Bischoff E、K¨opf F 和 Hohmann S（2019）基于最大熵逆强化学习的逆动态博弈。*arXiv*预印本 *arXiv:1911.07503*。JuliaGameTheoreticPlanning团队及贡献者（2023）MCPTrajectoryGameSolver.jl：Julia中的可微轨迹博弈求解器。
https://github.com/JuliaGameTheoreticPlanning/MCPTrajectoryGameSolver.jl。访问日期：2025-12-19。Kingma DP 和 Ba J（2015）Adam：一种用于随机优化的方法。载于：学习表征国际会议（*ICLR*）论文集。网址：https://arxiv.org/abs/1412.6980。Kingma DP 和 Welling M（2014）自编码变分贝叶斯。学习表征国际会议（*ICLR*）论文集。Laine F、Fridovich-Keil D、Chiu CY 和 Tomlin C（2023）近似广义反馈纳什均衡的计算。*SIAM*优化杂志33(1)：294–318。Le Cleac'h S、Schwager M 和 Manchester Z（2021）LUCIDGames：在线无迹逆动态博弈
用于自适应轨迹预测与规划。*IEEE*机器人与自动化快报（*RA-L*）6(3)：5485–5492. Lee N, Choi W, Vernaza P, Choy CB, Torr PH and Chandraker M (2017) Desire：在动态场景中对交互智能体进行远期预测。载于：*IEEE*计算机视觉与模式识别会议论文集。第336–345页。Li J, Chiu CY, Peters L, Palafox F, Karabag M, Alonso- Mora J, Sojoudi S, Tomlin C and Fridovich-Keil D (2023a) 情景博弈ADMM：一种用于随机非合作博弈的并行化情景求解器。载于：*2023*年第*62*届*IEEE*决策与控制会议（*CDC*）。IEEE，第8093–8099页。Li J, Chiu CY, Peters L, Sojoudi S, Tomlin CJ and Fridovich-Keil D (2023b) 基于含噪部分状态观测和不完整轨迹推断反馈动态博弈的成本。载于：国际自主智能体与多智能体系统会议论文集（*AAMAS*）。Lidard J, So O, Zhang Y, DeCastro J, Cui X, Huang X, Kuo YL, Leonard J, Balachandran A, Leonard N 等（2023）Nashformer：利用局部纳什均衡进行语义多样化轨迹预测。*arXiv*预印本 *arXiv:2305.17600*。Liu X, Peters L 和 Alonso-Mora J（2023）学习在轨迹博弈中与目标函数未知的对手进行博弈。*IEEE*机器人与自动化快报8(7): 4139–4146. DOI: 10.1109/LRA.2023.3280809。Liu X, Peters L, Alonso-Mora J, Topcu U 和 Fridovich- Keil D（2024）自动编码贝叶斯逆博弈。载于：机器人学算法基础国际研讨会（*WAFR*）。网址：https://arxiv.org/abs/2402.08902。Mehr N, Wang M, Bhatt M 和 Schwager M（2023）最大熵多智能体动态博弈：正向与逆向解。*IEEE*机器人学汇刊（*TRO*）39(3): 1801–1815. DOI:10.1109/TRO.2022.3232300。Meo C 和 Lanillos P（2021）多模态变分自编码器主动推理控制器。载于：*2021*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，第2693–2699页。Molloy TL, Inga Charaja J, Hohmann S 和 Perez T（2022）逆非合作动态博弈。载于：《逆最优控制与逆非合作动态博弈理论：极小值原理方法》。施普林格，第143–187页。Monderer D 和 Shapley LS（1996）势博弈。《博弈与经济行为》14(1)：124–143。Murphy KP（2022）《概率机器学习：导论》。麻省理工学院出版社。网址：probml.ai。Murphy KP（2023）《概率机器学习：高级专题》。麻省理工学院出版社。Peters L（2020）在人机交互的常和博弈中应对意图不确定性。硕士论文，汉堡工业大学。Peters L, Bajcsy A, Chiu CY, Fridovich-Keil D, Laine F, Ferranti L 和 Alonso-Mora J（2024）用于多智能体交互的应变博弈。*IEEE*机器人与自动化快报（*RA-L*）。Peters L, Rubies-Royo V, Tomlin CJ, Ferranti L, Alonso-Mora J, Stachniss C 和 Fridovich-Keil D（2023a）在动态博弈中根据部分观测在线与离线学习参与者目标函数。《国际机器人研究期刊》42(10): 917–937. DOI:10.1177/02783649231182453。网址：http://journals.

sagepub.com/doi/10.1177/02783649231182453。Peters L, Rubies-Royo V, Tomlin CJ, Ferranti L, Alonso-Mora J, Stachniss C 和 Fridovich-Keil D（2023b）在动态博弈中根据部分观测在线与离线学习参与者目标函数。《国际机器人研究期刊》（IJRR）42(10): 917–937。Rothfuß S, Inga J, K¨opf F, Flad M 和 Hohmann S（2017）非合作微分博弈中的逆最优控制与辨识。IFAC-PapersOnLine 50(1): 14909–14915. DOI:https://doi.org/10.1016/j.ifacol.2017.08.2538。Salzmann T, Ivanovic B, Chakravarty P 和 Pavone M（2020）Trajectron++：基于异构数据的动态可行轨迹预测。载于：欧洲计算机视觉会议。施普林格，第683–700页。Schonfeld E, Ebrahimi S, Sinha S, Darrell T 和 Akata Z（2019）通过对齐变分自编码器实现广义零样本与少样本学习。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第8247–8255页。Schwarting W, Pierson A, Alonso-Mora J, Karaman S 和 Rus D（2019）自动驾驶车辆的社会行为。《美国国家科学院院刊》116(50): 24972–24978。Shi Y, Siddharth N, Paige B 和 Torr PHS（2019）用于多模态深度生成模型的变分混合专家自编码器。《神经信息处理系统进展》（NeurIPS）论文集32。网址：https://arxiv.org/abs/1911.03393。Sim´eoni O, Vo HV, Seitzer M, Baldassarre F, Oquab M, Jose C, Khalidov V, Szafraniec M, Yi S, Ramamonjisoa M, Massa F, Haziza D, Wehrstedt L, Wang J, Darcet T, Moutakanni T, Sentana L, Roberts C, Vedaldi A, Tolan J, Brandt J, Couprie C, Mairal J, J´egou H, Labatut P 和 Bojanowski P（2025）DINOv3。arXiv预印本 arXiv:2508.10104 DOI:10.48550/arXiv.2508. 10104。网址：https://arxiv.org/abs/2508.10104。Spica R, Cristofalo E, Wang Z, Montijano E 和 Schwager M（2020）面向自动驾驶双人无人机竞速的实时博弈论规划器。IEEE机器人学汇刊（TRO）36(5): 1389–1403。Sun MM, Trautman P 和 Murphey T（2025）基于生成式轨迹模型的逆混合策略博弈。载于：IEEE机器人与自动化国际会议（ICRA）论文集。第8181–8188页。DOI:10.1109/ICRA55743.2025.11128615。Sutter TM, Daunhawer I 和 Vogt JE（2021）广义多模态ELBO。《国际学习表征会议》（ICLR）论文集。Suzuki M, Nakayama K 和 Matsuo Y（2017）基于深度生成模型的联合多模态学习。《国际学习表征会议》（ICLR）论文集。Takeishi N 和 Kalousis A（2021）用于鲁棒且可解释的生成建模的物理集成变分自编码器。载于：Beygelzimer A, Dauphin Y, Liang P 和 Vaughan JW（编）《神经信息处理系统进展》（NeurIPS）论文集。网址：https://openreview.net/ forum?id=0p0gt1Pn2Gv。Thoreau R, Risser L, Achard V, Berthelot B 和 Briottet X（2025）物理信息变分自编码器：提升对环境变化因素的鲁棒性。《机器学习》114(9)。DOI:10.1007/s10994-025-06829-7。网址：https: //doi.org/10.1007/s10994-025-06829-7。Wang M, Wang Z, Talbot J, Gerdes JC 和 Schwager M（2021）面向多车辆场景中自动驾驶汽车的博弈论规划

竞争场景。*IEEE*机器人学汇刊 *(TRO)* 37(4): 1313–1325. DOI:10.1109/TRO.2020.3047521. Wang Z, Spica R and Schwager M (2019) 多机器人竞速的博弈论运动规划。载于：分布式自主机器人系统：第*14*届国际研讨会。施普林格，pp. 225–238. Waugh K, Ziebart BD and Bagnell JA (2011) 计算合理化：逆均衡问题。 国际机器学习会议论文集 *(ICML)* 。Williams G, Goldfain B, Drews P, Rehg JM and Theodorou EA (2018) 用于自主地面车辆间敏捷交互的最佳响应模型预测控制。载于：*2018*年*IEEE*机器人与自动化国际会议 *(ICRA)*。Wu M and Goodman N (2018) 面向可扩展弱监督学习的多模态生成模型。神经信息处理系统进展论文集 *(NeurIPS)* 31. Yuan Y, WengX, Ou Y andKitani KM (2021) Agentformer：用于社会时序多智能体预测的智能体感知型 Transformer。载于： *IEEE/CVF*国际计算机视觉会议论文集。pp. 9813–9823. Zambelli M, Cully A and Demiris Y (2020) 基于部分信息进行预测和控制的多模态表征模型。 机器人与自主系统123: 103312. Zhong W and Meidani H (2023) Pi-vae：用于随机微分方程的物理信息变分自编码器。计算力学与工程方法403:115664. DOI:https://doi.org/10.1016/j.cma.2022.115664.URL
h
t
t
p
s://
w
w
w.
s
c
i
e
ncedirect.com/science/article/pii/S0045782522006193。Zhou K, Chen C, Wang B, Saputra MRU, Trigoni N and Markham A (2021) Vmloc：用于基于学习的多模态相机定位的变分融合。载于：*AAAI*人工智能会议论文集，第35卷。pp. 6165–6173. Zhu EL and Borrelli F (2023) 一种用于求解开环广义纳什均衡的序列二次规划方法。载于： *2023*年*IEEE*机器人与自动化国际会议 *(ICRA)*。IEEE，pp. 3211–3217。