2 × 2 博弈中的逆向学习：从合成交互到交通仿真

Daniela Aguirre Salazar、Firas Moatemri 和 Tatiana Tatarenko

arXiv:2601.10367v1 [cs.GT] 15 Jan 2026

摘要——**从有限的行为数据中理解智能体如何协调或竞争，是对交通、机器人学及其他多智能体系统中的策略交互进行建模的核心。在本研究中，我们探讨逆博弈论学习的以下两种互补表述：（i）一种针对 2×2 博弈的**闭式相关均衡极大似然**估计器（CE-ML）；以及（ii）一种*Logit*最佳回应极大似然估计器（LBR-ML），通过随机响应过程捕捉长期适应动态。这两种方法共同涵盖了静态均衡一致性与动态行为真实性之间的各种情形。我们在合成的“胆小鬼博弈”和使用 SUMO 仿真的交通交互场景中对其进行评估，比较参数恢复效果和分布拟合结果。结果揭示了不同模型在可解释性、计算可处理性和行为表达能力之间存在明显的权衡。**

I. 引言

博弈论为建模多个决策智能体之间的策略交互提供了原则性基础。传统方法假设每个玩家的效用函数已知，并据此预测均衡行为。相比之下，逆向视角试图从观测到的行为中推断这些潜在效用，解释智能体为何采取相应行动，而不仅仅是描述由此产生的均衡。这种逆博弈论学习范式尤其适用于交通、机器人学和人机交互等领域，在这些领域中，偏好和协调机制是隐含的，无法直接测量 [1], [2], [3]。

逆向学习面临的一个核心挑战，是选择最能捕捉观测到的协调行为或均衡概念。经典的纳什均衡（NE）假设策略选择完全独立且具有完全理性，但它在一般博弈中通常难以计算，属于 PPAD 完全 [4], 问题，并且往往与经验数据不一致。在许多多智能体场景中，包括驾驶场景，智能体会表现出部分协调或相关行为，偏离严格的纳什预测 [5], [6]。

为解决这些局限，相关均衡（CE） [7] 兼具易于计算 [8] 和更贴近真实行为的优点。CE 放宽了纳什均衡的独立性假设，允许智能体的行动通过共享信号或习得的预期 [9]呈条件相关。此外，已知基于遗憾的适应性动态会收敛至相关均衡集 [10], ，从而提供了一种在均衡理论与重复博弈学习之间建立自然桥梁，这对于交通等领域尤为重要，因为驾驶员会不断调整自身行为以适应彼此的行为。在此基础上，近期研究探索了利用逆向方法，根据观测到的相关博弈行为估计与之相符的潜在效用。其中，逆向相关均衡（*ICE*）框架 [11] 提供了一种有理有据的基线：它推断出一组效用，使观测到的经验分布成为近似相关均衡。然而， [11] 中基于 ICE 的效用学习回避了对潜在分布进行有原则的统计估计。所提出的线性规划方法仅在经验分布下按期望值施加相关均衡约束，而没有对相关均衡分布本身进行建模和拟合。

另一种旨在刻画更真实策略行为的解概念是量化响应均衡（*QRE*） [12], ，它通过允许参与者做出概率性选择，而非完全理性的选择，来建模有限理性。逆量化响应均衡框架已被用于分析交通领域中的冲突交互，这类交互通常表示为 2×2 标准式博弈 [13], [14]。然而，这些研究并未为 QRE 所隐含的联合行动分布的具体形式提供理论依据。此外，计算 QRE 与计算纳什均衡 [12] 同样困难，因此属于 PPAD 完全问题，这使得逆量化响应均衡学习通常难以处理。

为解决这些局限，我们提出两种新颖的逆向学习公式，以拓展 2×2标准式博弈中基于均衡的推断的适用范围并增强其可解释性。这种结构常见于交通交互、安全场景以及人机决策等实际领域 [13], [15], [16]。第一种方法是闭式相关均衡最大似然（CE-ML）估计器，它利用 2×2博弈 [17]中相关均衡多胞体的解析几何性质，直接估计均衡分布并高效进行参数拟合。第二种方法是*Logit*最佳回应最大似然（LBR-ML）框架，它不再采用静态均衡推理，而是将博弈的长期分布建模为同步 Logit最佳响应动态 [18]的平稳状态。这两种公式共同涵盖了一系列建模选择，从基于期望的均衡一致性（ICE），到解析可处理估计（CE-ML），再到具有有限理性的、基于动态且贴近行为现实的建模（LBR-ML）。

我们在两个

作者供职于德国达姆施塔特工业大学控制理论与智能系统系。电子邮件：daniela.aguirre@tu-darmstadt.de、firas.moatemri@stud.tu-darmstadt.de、tatiana.tatarenko@tu-darmstadt.de。

具有代表性的领域中对这三种方法进行了经验评估。首先，在真实效用已知的合成“胆小鬼博弈”中，我们测量参数恢复准确率和均衡一致性。其次，在基于 SUMO 的交通仿真中，我们考察推断出的效用是否能够在无信号交叉口中复现符合现实的让行与通行模式。这项评估展示了建模交互系统中策略行为的方法差异及其实际影响。

II. 问题表述与均衡概念

A. 逆博弈论学习

我们考虑一个有限正规型博弈 $\Gamma_N^w = \Gamma\{N, \{A_i\}, \{u_i^{w_i}\}\}$，参与者来自集合 $[N] = \{1, \dots, N\}$。每位参与者 $i \in [N]$ 都有一个行动集 A_i，联合行动集为 $A = \prod_{i\in[N]} A_i$。我们用 n 表示集合 A 的基数。对于联合行动 $a = (a_i, a_{-i}) \in A$，a_i 表示参与者 i 的行动，而 a_{-i} 表示除 i 之外所有参与者的行动。

每位参与者获得效用 $u_i^{w_i}(a) \in \mathbb{R}$，并假设该效用关于已知的特征表示 $\phi_i(a) \in \mathbb{R}^{d_i}$ 呈线性：

$$u_i^{w_i}(a) = \phi_i(a)^\top w_i, \tag{1}$$

其中 $w_i \in \mathbb{R}^{d_i}$ 是待估计的参与者特定权重。效用表达式（1）通过特征 $\phi_i(a)$ 及其相对重要性（由 w_i 编码），刻画了决策中的可解释因素，例如风险、舒适度相关成本或社会偏好。经典分析假设效用 $\{u_i^{w_i}\}$ 已知，并旨在预测 Γ_N^w 中的稳定状态（即所谓的均衡）。**逆博弈论学习**则反转了这一逻辑：给定观测到的联合行动 $\mathcal{D} = \{a^{(t)}\}_{t=1}^T$，我们寻求效用参数 $w = (w_1, \dots, w_N) \in \mathbb{R}^{Nd}$，使观测结果与该博弈中的某个确定均衡策略相一致。

下一节将介绍几种适用于有限标准式博弈的均衡概念。

B. 均衡概念

博弈 Γ_N^w 中玩家 i 的混合策略是一个概率分布 $\sigma_i \in \Delta(A_i)$，其中 $\Delta(A_i)$ 是 A_i 上所有可能概率分布的集合，而联合混合策略为 $\sigma = (\sigma_i)_{i\in[N]} \in \Delta(A) = \prod_{i\in[N]} \Delta(A_i)$。除非另有说明，假设 σ 是一个维度等于 A 基数的行向量 。因此，玩家 i 在 σ 下的期望效用定义为

$$\tilde{u}_i^{w_i}(\sigma) = \sum_{a\in A} \sigma[a]\, u_i^{w_i}(a), \tag{2}$$

其中 $\sigma[a] = \prod_{j\in[N]} \sigma_j[a_j]$ 表示联合混合策略赋予纯行动组合 $a = (a_1, \dots, a_N)$ 的概率。本文中，我们用符号 $\sigma[\cdot]$ 表示特定联合行动的概率，而 $\sigma(\cdot)$ 表示策略组合对模型／博弈参数的参数依赖关系，例如待估计的向量 w 。当这两个方面同时出现时，我们写作 $\sigma(\cdot)[\cdot]$.[1]

经典的混合策略纳什均衡 (NE)要求每个玩家的混合策略都是对其他玩家策略的最佳回应。形式上，对于所有 i 和所有$\sigma_i \in \Delta(A_i)$，如果 $\sigma^\star_{\mathrm{NE}}(w)$ 是一个 NE，则

$$\tilde{u}_i^{w_i}\big(\sigma^\star_{\mathrm{NE},i}(w),\ \sigma^\star_{\mathrm{NE},-i}(w)\big) \ \geq\ \tilde{u}_i^{w_i}\big(\sigma_i,\ \sigma^\star_{\mathrm{NE},-i}(w)\big).^{2}$$

一般而言，计算 NE 是 PPAD 完全的 [4], , 且实证研究表明，人类的行动往往违反独立性假设 [5], [6], , 这促使人们提出允许相关或自适应协调的均衡概念。本文聚焦于这类概念中的两个例子：相关均衡 和*Logit*最佳响应动态的平稳状态。

a) 概念一：相关均衡：相关均衡(CE) [7]通过允许玩家的行动之间存在相关性，并以联合分布表示（通常由引入系统的外部相关装置生成），推广了纳什均衡。分布 $\hat{}\sigma^\star(w)$ 在 Γ_N^w 中是相关均衡，当且仅当对所有玩家 i 及所有偏离行动 $a_i' \in A_i$，

$$\sum_{a\in A} \hat{\sigma}^\star(w)[a_i, a_{-i}] \left[u_i^{w_i}(a_i, a_{-i}) - u_i^{w_i}(a_i', a_{-i})\right] \geq 0. \tag{3}$$

相关均衡保留了凸性，并且可通过线性规划高效求解；相关装置则提供了一种概率机制，据此可以从数据中估计均衡联合行动。

b) 概念二：*Logit*最佳回应动态的平稳状态：或者，基于*Logit*最佳回应(LBR) [18] 的行为模型通过引入由玩家特定理性参数 $\lambda_i \geq 0$支配的随机选择，放宽完全理性和集中式相关的假设。令 $\lambda = (\lambda_1、\ .....、\ \lambda_N)$表示理性水平向量。将联合行动集枚举为$A = \{a(1)、\ .....、\ a(n)\}$，并令 $a_i(l)$表示 i号玩家在 $a(l)$中的分量。给定当前联合行动 $a(k) \in A$，LBR诱导的混合策略为

$$\sigma(\lambda, w)[a(l)|a(k)] = \prod_{i=1}^{N} \sigma_i(\lambda_i, w_i)[a_i(l)|a_{-i}(k)], \qquad a(l) \in A,$$

$$\sigma_i(\lambda_i, w_i)[a_i(l)|a_{-i}(k)] = \frac{\exp\big(\lambda_i\, u_i^{w_i}(a_i(l), a_{-i}(k))\big)}{\sum_{a_i'\in A_i} \exp\big(\lambda_i\, u_i^{w_i}(a_i', a_{-i}(k))\big)}. \tag{4}$$

较高的 λ_i 表示玩家 i的行为更具确定性（即更“理性”）。当 λ_i =为 0 时， A_i上的各项选择概率相等；而随着 $\lambda_i \to \infty$增大，$\sigma_i(\lambda_i,\ w_i)[\cdot|a_{-i}(k)]$ 会集中于对 $a_{-i}(k)$的最佳回应（假设最佳回应唯一，或采用固定的平局处理规则）。

[1]例如， $\sigma(w)[a]$ 表示在效用函数参数化 w下，混合策略 σ 中行动 a 的概率。

[2]我们用 σ_{-i} 表示除玩家 i 之外所有玩家的联合混合策略

值得注意的是，遵循 LBR 并不需要协调器或相关装置，而 CE 概念则需要。每位玩家都需要观察上一次实现的联合行动 $a(k)$，并了解相应的收益$u_i^{w_i}(\cdot,\ a_{-i}(k))$. 因此，决策仅取决于最近一次联合行动，这使得该过程具有短视的 和无记忆的 [18], [20]特性。这使当前的 LBR 动态区别于量化响应均衡 (QRE)：在后者中，Logit响应是针对混合策略下的期望效用作出的，因此需要玩家相互交换各自的混合策略。

如果参与者在每个阶段同时遵循LBR规则，则在有限集合 A 上诱导出的过程是一个齐次马尔可夫链，其转移矩阵为

$$P(\lambda, w) = (p_{kl}(\lambda, w))_{k,l=1}^{n}, \tag{5}$$

其中， $p_{kl}(\lambda, w) = \sigma(\lambda, w)[a(k)|a(l)]$ 由式（4）定义。也就是说， $p_{kl}(\lambda, w)$是从联合行动 $a(k)$转移到 $a(l)$的概率。令$\sigma^t \in \Delta(A)$表示时刻 t联合行动上的分布。那么，联合混合策略的演化为

$$\sigma^{t+1} = \sigma^t\, P(\lambda, w), \tag{6}$$

给定参与者从某个初始 $\sigma^0 \in \Delta(A)$出发。

记 $\sigma^\star(\lambda, w)$为 $P(\lambda, w)$的平稳分布，即

$$\sigma^\star(\lambda, w)\, P(\lambda, w) = \sigma^\star(\lambda, w). \tag{7}$$

由于 $P(\lambda, w)$具有行随机性，平稳分布必然存在。

如果对所有 i 都有 $0 \le \lambda_i < \infty$ 且效用是有限的，则式（4）中的每一项都严格为正，因此 $P(\lambda, w)$是一个本原随机矩阵。根据有限马尔可夫链的Perron-Frobenius定理， $P(\lambda, w)$存在唯一的平稳分布，并且该链以指数速度收敛于该分布 [21]。因此，我们得到以下结果。

定理*1*： 设 Γ_N^w 中的参与者根据式（6）随时间同时选择 LBR 策略，理性水平 $\lambda_i < \infty$ 有限，并从任意混合策略$\sigma^0 \in \Delta(A)$开始。则存在唯一的平稳分布$\sigma^\star(\lambda, w)$，使得 $\lim_{t\to\infty} \sigma^t = \sigma^\star(\lambda, w)$。

上述定理意味着，该动态会稳定在$P(\lambda, w)$的唯一平稳分布 $\sigma^\star(\lambda, w)$附近。此 $\sigma^\star(\lambda, w)$称为 Γ_N^w中的*Logit*最佳回应动态的平稳状态，并作为我们第二个逆博弈论学习框架中的参考均衡分布。

III. 2×2 博弈中的逆博弈论学习方法

本节将详细介绍两种逆博弈论学习框架：(i) 针对 2×2 博弈的闭式相关均衡估计器（CE-ML），以及 (ii) 基于 Logit 响应动态平稳分布的*Logit* 最佳响应学习（LBR-ML）方法。每种方法都采用上一节中的特定均衡解概念，将观测到的行为数据与按式 (1) 参数化的潜在效用联系起来。

具体而言，考虑博弈Γ_2^w，其中$[2] = \{1, 2\}$ 为参与者集合，玩家 1 为行玩家，玩家 2 为列玩家（见表 I）。对于$i \in [2]$，令 $A_i = \{a_i^1, a_i^2\}$为玩家 i的行动集。联合行动集为 $A = A_1 \times A_2$，枚举如下： $A = \{a(1) = (a_1^1, a_2^1)\ \ a(2) = (a_1^1, a_2^2)\ \ a(3) = (a_1^2, a_2^1)\ \ a(4) = (a_1^2, a_2^2)\}$. 对于每位玩家 i，联合行动 $a(l)$下的收益记为 $u_i^{w_i}(a(l))$。与之前一样，$w_1 \in \mathbb{R}^d$ 和 $w_2 \in \mathbb{R}^d$ 是待估计的效用参数向量（见式（1）中的线性效用设定）。

表 I 双人 2×2 博弈：（A）收益矩阵（$u_1^{w_i},\ u_2^{w_i}$）。

	a_2^1	a_2^2
a_1^1	$(u_1^{w_1}(a(1))\ \ u_2^{w_2}(a(1)))$	$(u_1^{w_1}(a(2))\ \ u_2^{w_2}(a(2)))$
a_1^2	$(u_1^{w_1}(a(3))\ \ u_2^{w_2}(a(3)))$	$(u_1^{w_1}(a(4))\ \ u_2^{w_2}(a(4)))$

A. 方法一： *2×2 CE* 最大似然闭式解（*CE-ML*）

文献 [17]已证明，对于 2×2 标准式博弈这一类博弈，相关均衡（CE）集可以用由四个线性不等式 (3) 和单纯形约束定义的紧凑闭式特征来描述。从几何上看，该可行域构成一个凸多胞体，可表示为少数极点的凸包，这些极点对应可行的纯策略或混合策略联合行动配置。

我们采用文献 [17],提出的以下假设，该假设使得能够完整刻画 2×2标准式博弈中的相关均衡集。

假设（不存在被支配策略）。任一玩家的行动均不被严格支配；特别是，$u_1^{w_1}(a(1)) \neq u_1^{w_1}(a(3))$， $u_1^{w_1}(a(4)) \neq u_1^{w_1}(a(2))$， $u_2^{w_2}(a(1)) \neq u_2^{w_2}(a(2))$， $u_2^{w_2}(a(4)) \neq u_2^{w_2}a((3))$.

根据表 I 中的收益矩阵 $(u_1^{w_1}(a(l))\ \ u_2^{w_2}(a(l)))_{l\in\{1,2,3,4\}}$ 并考虑上述假设，我们定义

$$\alpha^{w_1} = \frac{|u_1^{w_1}(a(1)) - u_1^{w_1}(a(3))|}{|u_1^{w_1}(a(4)) - u_1^{w_1}(a(2))|},$$
$$\beta^{w_2} = \frac{|u_2^{w_2}(a(1)) - u_2^{w_2}(a(2))|}{|u_2^{w_2}(a(4)) - u_2^{w_2}(a(3))|}.$$

定理 *2*：考虑一个（严格）协调博弈 Γ^w，其中 $u_1^{w_1}(a(1)) > u_1^{w_1}(a(3))$, $u_1^{w_1}(a(4)) > u_1^{w_1}(a(2))$,$u_2^{w_2}(a(1)) > u_2^{w_2}(a(2))$, $u_2^{w_2}(a(4)) > u_2^{w_2}(a(3))$。那么， Γ_2^w 中的相关均衡集合是一个具有五个顶点 $\{\hat{}\sigma_{(v)}^\star(w)\}_{v=1}^5$的多胞体，如表 II [3]所列，其中我们将联合行动 $a(l)$，$l = 1$，2，3，4 的概率简记为 $\hat{}\sigma^\star(w)[a(l)]$。
证明见 [17] 。

[3]为简化记号，在表 II 中，我们分别省略了 w_1 和 w_2 对 α 和 β 的依赖关系。

表 II 严格协调情形下相关均衡多胞体的五个顶点，如 [17]中所述。

σ	$\hat{\sigma}^\star(w)[a(1)]$	$\hat{\sigma}^\star(w)[a(4)]$	$\hat{\sigma}^\star(w)[a(3)]$	$\hat{\sigma}^\star(w)[a(2)]$
$\hat{\sigma}^\star_{(1)}(w)$	1	0	0	0
$\hat{\sigma}^\star_{(2)}(w)$	0	1	0	0
$\hat{\sigma}^\star_{(3)}(w)$	$\frac{1}{(1+\alpha)(1+\beta)}$	$\frac{\alpha\beta}{(1+\alpha)(1+\beta)}$	$\frac{\beta}{(1+\alpha)(1+\beta)}$	$\frac{\alpha}{(1+\alpha)(1+\beta)}$
$\hat{\sigma}^\star_{(4)}(w)$	$\frac{1}{1+\beta+\alpha\beta}$	$\frac{\alpha\beta}{1+\beta+\alpha\beta}$	$\frac{\beta}{1+\beta+\alpha\beta}$	0
$\hat{\sigma}^\star_{(5)}(w)$	$\frac{1}{1+\alpha+\alpha\beta}$	$\frac{\alpha\beta}{1+\alpha+\alpha\beta}$	0	$\frac{\alpha}{1+\alpha+\alpha\beta}$

注 *1*（见 *[17]*）：对于一个具有反向严格不等式 $u_1^{w_1}(a(1)) < u_1^{w_1}(a(3))$、$u_1^{w_1}(a(4)) < u_1^{w_1}(a(2))$、$u_2^{w_2}(a(1)) < u_2^{w_2}(a(2))$、$u_2^{w_2}(a(4)) < u_2^{w_2}(a(3))$ 的反协调博弈，通过将 β^{w_2} 替换为 $1/\beta^{w_2}$ 并交换非对角概率与对角概率，可以得到表 II 中的顶点。在支配情形下，相关均衡与纳什均衡的集合重合，并收缩为单个点。

在协调博弈 Γ_2^w 中，根据定理 2，相关均衡集合中的每个元素都可以表示为

$$\hat{\sigma}^\star(w,y)=\sum_{v=1}^{5} y_v \hat{\sigma}^\star_{(v)}(w), \tag{8}$$

对于某些 $y\in\Delta_5=\{(y_1,\ldots,y_5):\ y_j\ge 0,\ \sum_{j=1}^5 y_j=1\}$。

接下来，我们假设 Γ_2^w 中的参与者遵循某种相关均衡，并给定了 T 个关于相应联合行动的独立观测 $\mathcal{D}=\{a^{(t)}=(a_1^{(t)},\ a_2^{(t)})\in A\}_{t=1}^T$。

然后，使用式 (8) 中相关均衡的参数化表达式 $\hat{\sigma}^\star(w,y)=(\hat{\sigma}^\star(w,y)[a(1)],\hat{\sigma}^\star(w,y)[a(4)],\hat{\sigma}^\star(w,y)[a(3)],\hat{\sigma}^\star(w,y)[a(2)])$，我们计算每个观测中实际出现的行动组合的概率，并将它们累积为似然函数：$L_T(\prod_{t=1}^T \sigma^\star\ w\ y\ [a^{(t)}]\qquad \sigma^\star\ w\ y\ [a^{(t)}],)\hat{}$ (,)。在最后一个表达式中，如果观测 $a^{(t)}$ 是相应的联合行动 $a(l)$，那么 $\hat{}(,)$ 就等于 $\hat{}\sigma^\star(w,y)[a(l)]$，中的一个。最大似然估计求解以下问题：

$$\min_{w,y}\{-\log(L_T(w,y))=-\sum_{t=1}^{T}\log\big(\hat{\sigma}^\star(w,y)[a^{(t)}]\big)\},\quad \text{s.t. } y\in\Delta_5. \tag{9}$$

因此，上述优化问题的目标是找到最佳权重 y_v，$v=1,\ldots,5$，以及效用参数 w，使其能够通过均衡策略 $\hat{}\sigma^\star(w,\ y)$共同描述数据集 $\mathcal{D}$ 中的经验行动频率；也就是说，找出使观测到的行动在假定的博弈论行为模型下最有可能出现的参数配置。因此，式（9）的解 (w^*,y^*) 中的部分 w^*，在 Γ_2^w中解决了逆博弈论学习问题，前提是参与者选择相关均衡作为其策略互动的解概念。

B. 方法二：*Logit* 最佳响应学习（*LBR-ML*）

上一小节介绍的基于相关均衡的表述（如 CE-ML）假设观测到的行为受到相关装置的影响，从而生成相应的稳定分布。相比之下，本小节中的 LBR-ML 方法对在重复互动中不断相互调整的有限理性主体进行建模，并通过第 II 节介绍的对数最佳响应动态的平稳状态将其形式化。

我们再次考虑博弈 $2\times 2\ \Gamma_2^w$（见表 I），其中有 $A_i=\{a_i^1,a_i^2\}$、$i=1$、2，联合行动集为 $A=A_1\times A_2$，与之前一样，其定义如下：$A=\{a(1)=(a_1^1,\ a_2^1)\ a(2)=(a_1^1,\ a_2^2)\ a(3)=(a_1^2,\ a_2^1)\ a(4)=(a_1^2,\ a_2^2)\}$。现在假设参与者同时遵循有限正值的 Logit最佳响应动态，其中 λ_1 和λ_2（见式（4））。因此，所诱导的过程是一个具有转移矩阵 $P(\lambda,\ w)=(p_{kl}(\lambda,\ w))_{k,\ l=1}^4$的四状态马尔可夫链。根据定理1，对于有限的 λ_i，该动态收敛到联合行动集 A 上的唯一平稳分布 $\sigma^\star(\lambda,\ w)$，且该分布满足式（7）。在 LBR-ML 方法中，假定该分布为均衡解概念。接下来，我们对所考虑博弈 Γ_2^w 中的 $\sigma^\star(\lambda,\ w)$进行估计。

使用 Γ_2^w 中的式（4）定义单步逻辑斯蒂响应概率：

$$\begin{aligned}
s_1(\lambda_1,w_1)&=\sigma_1(\lambda_1,w_1)[a_1^1|a_2^1],\\
s_2(\lambda_1,w_1)&=\sigma_1(\lambda_1,w_1)[a_1^1|a_2^2],\\
t_1(\lambda_2,w_2)&=\sigma_2(\lambda_2,w_2)[a_2^1|a_1^1],\\
t_2(\lambda_2,w_2)&=\sigma_2(\lambda_2,w_2)[a_2^1|a_1^2].
\end{aligned}$$

于是，转移矩阵定义如下[4]：

$$\begin{aligned}
P(\lambda,w)&=(p_{kl}(\lambda,w))_{k,l=1}^4\\
&=\begin{bmatrix}
s_1t_1 & s_1(1-t_1) & (1-s_1)t_1 & (1-s_1)(1-t_1)\\
s_2t_1 & s_2(1-t_1) & (1-s_2)t_1 & (1-s_2)(1-t_1)\\
s_1t_2 & s_1(1-t_2) & (1-s_1)t_2 & (1-s_1)(1-t_2)\\
s_2t_2 & s_2(1-t_2) & (1-s_2)t_2 & (1-s_2)(1-t_2)
\end{bmatrix}.
\end{aligned} \tag{10}$$

定理1说明该系统的解存在且唯一

$$\sigma(\lambda,w)\,P(\lambda,w)=\sigma(\lambda,w),\quad \text{s.t. } \sigma(\lambda,w)\mathbf{1}=1.$$

对于所考虑的 2×2 博弈 Γ_2^w，上述线性方程组的闭式行向量解，即平稳分布 $\sigma^\star(\lambda,\ w)$，可按如下方式求得。根据平稳性，我们得到关于参与者行动边缘分布的以下仿射方程组：

$$\begin{aligned}
\sigma_1^\star(\lambda,w)[a_1^1]&=\sigma_2^\star(\lambda,w)[a_2^1]\,s_1+(1-\sigma_2^\star(\lambda,w)[a_2^1])s_2\\
\sigma_2^\star(\lambda,w)[a_2^1]&=\sigma_1^\star(\lambda,w)[a_1^1]\,t_1+(1-\sigma_1^\star(\lambda,w)[a_1^1])t_2.
\end{aligned}$$

通过根据上述不等式表示出 $\sigma_1^\star(\lambda,\ w)[a_1^1]$和 $\sigma_1^\star(\lambda,\ w)[a_2^1]$，我们得到 $\sigma_1^\star(\lambda,\ w)[a_1^1]w=\ \ s_2+\ (\ \ s_1-s_2-t_1-t_2)\ \ (s_1-s_2)\ \ (t_1-t_2)$ 以及$-\sigma_2^\star(\lambda,\ s_2\ \ t_1-t_2\neq 0\qquad 1-\qquad\qquad\qquad

标准式博弈，因此最终的平稳分布可以由边缘分布按如下方式构成：

$$\begin{aligned}\sigma^*(\lambda,w)[a(1)] &= xy,\\ \sigma^*(\lambda,w)[a(2)] &= x(1-y),\\ \sigma^*(\lambda,w)[a(3)] &= (1-x)y,\\ \sigma^*(\lambda,w)[a(4)] &= (1-x)(1-y),\end{aligned} \tag{11}$$

其中我们使用记号 $x = \sigma_1^\star(\lambda, w)[a_1^1]$, $y = \sigma_2^\star(\lambda, w)[a_2^1]$。

现在假设 Γ_2^w 中的参与者 Logit最佳回应动态已经稳定，并且已提供相应联合行动的 T 独立观测 $\mathcal{D} = \{a^{(t)} \in A\}_{t=1}^T$。然后，与上一小节类似，我们可以根据式（11）中的表达式计算每个观测对应的已实现行动组合的概率。因此，累积似然函数如下：$L_T(\lambda, w) = \prod_{t=1}^T \sigma^\star(\lambda, w)[a^{(t)}]$。在上式中，如果观测 $a^{(t)}$ 是相应的联合行动 $a(l)$，则 $\sigma^\star(\lambda, w)[a(l)]$，等于 $\sigma^\star(\lambda, w)[a(l)]$,。最大似然估计通过求解以下问题得到：

$$\begin{aligned}\min_{\lambda,w}\{-L_T(\lambda,w) = -\sum_{t=1}^{T}\log\big(\sigma^\star(\lambda,w)[a^{(t)}]\big)\},\\ \text{s.t. } \lambda_1,\lambda_2 \geq 0.\end{aligned} \tag{12}$$

因此，上述优化问题的目标是寻找最佳的参与者理性参数 $\lambda = (\lambda_1, \lambda_2)$和效用权重 w，使其通过平稳均衡分布 $\sigma^\star(\lambda, w)$共同复现数据集 $\mathcal{D}$中观察到的经验联合行动频率。该问题旨在识别出在对数最优响应动态下，使观察到的行为最有可能出现的参数。具体而言，式 (12) 的解 w^* 部分(λ^*, w^*)**解决了** Γ_2^w中的逆博弈论学习问题，因为参与者选择 LBR 动态的平稳分布作为其策略互动的解概念。

IV. 实验设置

我们在一个合成 2×2 博弈和一个 SUMO 交通场景上评估我们的方法，从受控条件逐步过渡到现实条件。在每种情况下，我们都指定一个真实博弈，通过 CE 或仿真协议生成数据，并拟合 ICE、CE-ML 和 LBR-ML，同时报告参数恢复结果和分布拟合结果。

A. 实验领域

1）合成“斗鸡博弈”：我们采用经典的 2×2斗鸡博弈互动，其中每位玩家选择前进或等待。效用遵循式（1）中的线性规格，使用来自 [11] 的特征映射和幅值；为简洁起见，此处省略特征定义。为设定真实值，我们将玩家的权重固定为本实验中定义的数值，在由此得到的博弈中计算最大熵相关均衡（有关这一相关均衡概念的更多详情，参见 [22] ），并从该相关均衡中抽取独立同分布的联合行动。这样便能确保数据与已知博弈的相关均衡一致，从而单独考察逆向步骤。

2）*SUMO* 中的交通互动博弈：为评估其现实性和适用性，我们在微观交通模拟器（SUMO[19]）中考察相同的 2×2 互动结构。该环境在抽象的“斗鸡博弈”设定基础上，引入了连续运动学依赖关系和随机驾驶员行为，为所提出的逆向学习模型提供了一个逼真的验证场景。

a) 场景设计：两个智能体（A 和 B）从正交方向驶向无信号交叉路口，并选择 $A_i = \{$等待、前进$\}$。令 v_i 表示速度，d_i 表示距冲突点的距离，$\tau_i = d_i/v_i$ 表示到达交叉口的时间，$\delta = \tau_2 - \tau_1$。图1展示了用于提取（v_i、d_i）并推导（τ_i、δ）的几何关系。

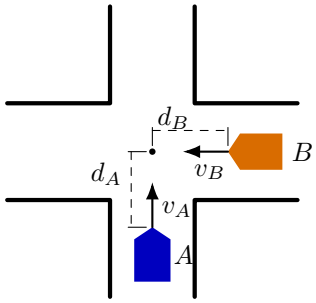


图1. SUMO 中使用的无信号交叉路口。距离 d_A、d_B 和速度 v_A、v_B 定义了时间 $\tau_i = d_i/v_i$ 和偏移量 $\delta = \tau_2 - \tau_1$。

b) 运动学相关效用：效用 $u_i^{w_i}(a)$ 遵循式（1）中的线性模型，其中，特征向量$\phi_i(a;\ \tau_i、\ \tau_{-i}、\ \delta)$ 对可解释的运动学量进行编码，以刻画紧迫性（τ_i^{-1}）、时间压力（τ_i）、近碰撞风险（$(|\delta|+\varepsilon)^{-1}$）和基准偏好；由此得到的收益表如下：

	a_2^1	a_2^2
a_1^1	$(-w_{1,7}\lvert\delta\rvert+w_{1,8},\ -w_{2,7}\lvert\delta\rvert+w_{2,8})$	$(-w_{1,3}\tau_1+w_{1,4},\ \frac{w_{2,1}}{\tau_2}+w_{2,2})$
a_1^2	$(\frac{w_{1,1}}{\tau_1}+w_{1,2},\ -w_{2,3}\tau_2+w_{2,4})$	$(-\frac{w_{1,5}}{\lvert\delta\rvert+\varepsilon}+w_{1,6},\ -\frac{w_{2,5}}{\lvert\delta\rvert+\varepsilon}+w_{2,6})$

(13)

请注意，引入 $\varepsilon = 10^{-6}$ 是为了避免潜在的除零问题。此设定遵循交通交互模型 [14] 中的风险/时间公式，以及无信号交叉口 [23]所采用的替代收益；另请参阅更广泛的自动驾驶汽车决策 [24]。

c) 数据集提取：我们在 SUMO [19]中遍历初始条件，并运行每个回合，直到智能体选择行动的决策点 $a_i \in \{$等待，前进$\}$。此时，我们记录 (v_i, d_i)，计算 (τ_i, δ)，并提取实际发生的联合行动 $a \in A_1 \times A_2$。每个数据点为$((a_1, a_2)\ \phi_1(a;\ \tau_1,\ \tau_2,\ \delta)\ \phi_2(a;\ \tau_2,\ \tau_1,\ \delta))$。

B. 比较框架

所有实验均比较以下逆向博弈论估计器：

- **ICE** [11]：一种凸规划，用于强制满足经验相关均衡一致性。
- **CE-ML**：用于拟合参数化相关均衡混合的最大似然估计器 $\hat{}\sigma^\star(w, y) =$

$\sum_{v=1}^{5} y_v \hat{\sigma}^{\star}_{(v)}(w)$，其中 $y \in \Delta_5$ 为顶点权重（参见优化问题（9））。

- **LBR-ML**：一种逻辑最佳响应模型，用于估计玩家效用 w 和理性参数（λ_1，λ_2），这些参数定义了诱导马尔可夫过程的平稳分布（见优化问题 (12)）。

C. 评估指标

我们报告两项互补的指标：

- **参数恢复：** 估计效用权重 $w_i^{\star}$ 与真实效用权重之间的平均绝对误差（MAE）和均方根误差（RMSE）。
- **分布拟合：** 相应均衡解中的经验联合行动概率与预测联合行动概率，以及决策准确率（正确预测行动的百分比）。

D. 实验设计

E1：通过相关均衡研究斗鸡博弈。

目的： 在完全已知的 2×2博弈中对每种方法进行基准测试，以验证其参数恢复能力和分布拟合效果。

设置： 使用合成的“斗鸡—胆小鬼”博弈，其真实权重为 $w_1^{\star} = [0.3, 0.7]^{\top}$ 和 $w_2^{\star} = [0.4, 0.6]^{\top}$。我们从该博弈的最大熵相关均衡中生成 $T \in \{500, 1000, 2000\}$ 个样本。

E2：交通（最大熵相关均衡真实值）。

目的： 评估当交通决策遵循精确相关均衡分布时的学习表现。

设置： 使用基于 SUMO 的交通博弈，其真实效用参数 $w^{\star}$ 取自最大熵相关均衡：

$$w_1^{\star} = [0.05,\ 0.00,\ 0.02,\ 0.00,\ 0.90,\ 0.00,\ 0.03,\ 0.00],$$
$$w_2^{\star} = [0.01,\ 0.00,\ 0.04,\ 0.00,\ 0.94,\ 0.00,\ 0.01,\ 0.00]. \tag{14}$$

E3：带信号相关性的交通。

目的： 考察相关装置产生部分协调行为时的稳健性。

设置： 信号选择联合行动 $(a_1, a_2) \sim \sigma_{\mathrm{CE}}$，并将建议分别私下发送给每位驾驶员，驾驶员按照分配给自己的行动行事。

E4：无协调交通。

目的： 在无法保证满足任何均衡结构的数据上测试这些模型。

设置： 在 SUMO 中模拟驾驶员之间的互动，不使用任何相关机制；驾驶员的随机差异性会引入行为噪声。

V. 结果

这些数据集提供了一个受控基准，用于评估各框架恢复真实效用参数 $w_i^{\star}$ 并在四项实验（E1-E4）中重现观测到的联合行动分布的能力。分别通过CE-ML和LBR-ML拟合得到的非权重参数 y 和 λ列于附录中的表 VI。

E1. 通过相关均衡实现的斗鸡博弈

数据集 $\mathcal{D}_{\mathrm{E1}} = \{a^{(t)}\}_{t=1}^{T}$ 提供了经验分布 σ_{CE}

所有方法，即 ICE、CE-ML 和 LBR-ML，均使用该方法通过最大似然估计参数 $w_i^{\star}$。

表 III 和图 2 总结了参数恢复和预测行动概率。

随着样本量 T 增加，尽管 ICE 和 LBR-ML 的表现仅有有限提升，CE-ML 的准确率和分布拟合却有明显改善，与真实 CE 最为接近，表现优于其他方法。对于 LBR-ML，这一趋势与拟合得到的理性参数 λ_i（见附录中的表 VI）有关，该参数在T=1000时骤升。较高的 λ 值意味着博弈行为更加确定，从而减少随机平滑，并增加对抽样噪声的敏感性，这与该时间跨度下误差暂时上升的现象一致。随着 T 进一步增大，λ值再次下降，使行为更加平滑，拟合效果也有所改善。

表 III 不同样本量 T下所有权重的总体 MAE/RMSE。数值越低越好；粗体表示每个 T下最佳的 MAE。

模型	$T = 500$	$T = 1000$	$T = 2000$
ICE	0.133 / 0.139	0.133 / 0.137	0.126 / 0.131
CE-ML	**0.133 / 0.138**	**0.114 / 0.114**	**0.084 / 0.097**
LBR-ML	0.155 / 0.194	0.433 / 0.442	0.131 / 0.176

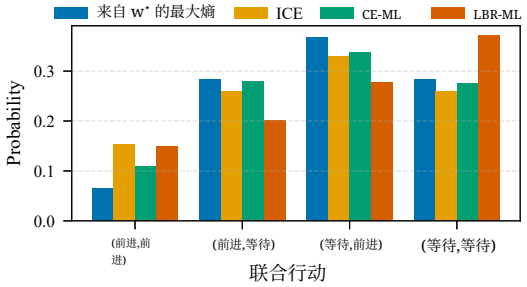


图2 。每个联合行动的均衡概率分布，T = 2000。

E2：交通（最大熵相关均衡真实值）

本实验使用从真实参数下的最大熵相关均衡中采样得到的数据集 $\mathcal{D}_{\mathrm{E2}} = \{a^{(t)}\}_{t=1}^{T}$，评估逆向模型在基于 SUMO 的交通博弈上的表现。

参数恢复误差见表 IV。CE-ML 的整体表现最佳，其次是 LBR-ML，然后是 ICE。考虑到 ICE 假设数据与 CE 一致，而 LBR-ML 依赖于不同的均衡概念，这是一个有趣的结果。对于 LBR-ML，拟合得到的理性参数 λ_i 在不同样本量下始终保持较高水平（见附录中的表 VI），这表明两个智能体的行为都近乎确定。这种稳定性说明，交互已经接近协调均衡结构，此时增加 T 几乎不会为模型带来新的随机信息供其学习。

表 IV 在所有权重下相对于样本量 T 的总体平均绝对误差/均方根误差。数值越低越好；粗体表示每个 T 下最佳的平均绝对误差。

模型	$T = 500$	$T = 1000$	$T = 2000$
ICE	0.111 / 0.143	0.111 / 0.144	0.111 / 0.144
CE-ML	**0.019 / 0.023**	**0.017 / 0.022**	**0.017 / 0.022**
LBR-ML	0.041 / 0.050	0.040 / 0.049	0.043 / 0.051

所考虑的均衡中的相应联合行动分布见图3。在这三个框架中，CE-ML 得出的分布与真实值最为一致。

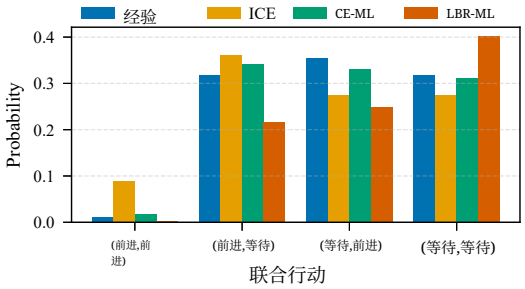


图3。每个联合行动对应的均衡概率分布，T = 2000。

$\mathcal{D}_{\text{E2}}$ 的结果证实，当观测数据遵循真实的相关均衡时，所有逆向建模形式都能够恢复一致的效用。为了进一步评估它们在协调情况下的稳健性，下一个实验引入了一种调节驾驶员行为的信号机制。

E3：带信号相关性的交通

相关装置生成联合行动（a_1，a_2）$\sim\sigma_{\text{CE}}$，并向每位玩家私下推荐行动，玩家则遵循其收到的信号。由此产生的数据集 $\mathcal{D}_{\text{E3}} = \{a^{(t)}\}_{t=1}^{500}$完整记录了符合相关均衡（CE）的信号介导的交互。该实验比较了 ICE 和 CE-ML 在恢复底层效用以及再现诱导的相关行为方面的能力。

准确率结果汇总于表V中，相应的联合行动分布如图4所示。正如在符合相关均衡（CE）的观测下所预期的那样，CE-ML 与观测数据的一致性最高，且参数误差最低，而 ICE 的拟合效果相对较弱。

表V 不同 SUMO 配置下的准确率。粗体表示性能最佳的模型。

模型	准确率（%）
ICE	41.80%
CE-ML	**86.40%**

估计得到的 CE-ML 混合权重 y_i 在不同样本量下保持稳定，并集中于生成数据的相同均衡成分，这表明 CE-ML 能够恢复由信号装置诱导的结构相关模式（见附录中的表 VI）。

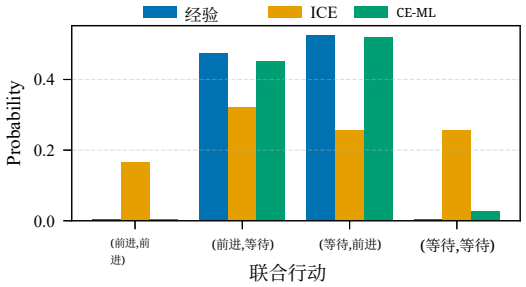


图4. 每种联合行动对应的均衡概率分布，T=500

E4：无协调交通

这项最后的实验考察了一种不存在明确协调或信号机制的情形。在 SUMO 仿真中，每位驾驶员独立行动，从而生成一个数据集 $\mathcal{D}_{\text{E4}} = \{a^{(t)}\}_{t=1}^{500}$，反映自然产生的、可能属于非均衡行为的情况。由于不存在相关装置，这种情形与.....尤其相关
评估模型在现实世界随机性和次优决策模式下的稳健性。

在这种情况下，由于均衡假设不再成立，CE-ML 和 ICE 已无法很好地匹配经验行为。因此，我们重点关注 LBR-ML，并比较固定 $\lambda_i = 1$ 的版本与从数据中估计 λ_i的变体（见附录中的表 VI）。两种配置的**准确率均为 72.6%**，但采用估计值的版本（λ_1=1.0 和 λ_2=3.0）生成的联合行动分布与实证频率更为接近（图 5）。这一结果表明，允许 λ_i 进行自适应，可以捕捉智能体的异质性有限理性，从而更灵活地表征非均衡交互，即使不存在协调或信号传递也是如此。

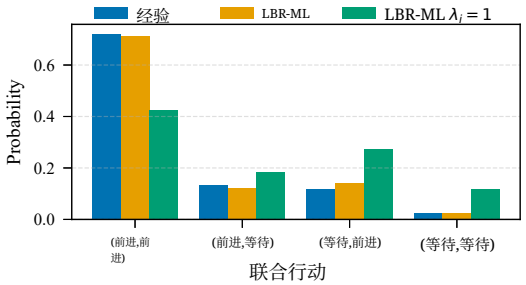


图 5. 各联合行动对应的均衡概率分布，T=500

在四项实验中，所提出的逆博弈论框架呈现出一致的趋势。当观测到的行为符合相关均衡假设时（E1–E3），CE-ML表现最佳；而在缺乏协调或存在噪声的条件下（E4），LBR-ML则更为稳健。ICE提供了一个通用的基线，但随着数据规模增加或偏离与相关均衡一致的设置，其改进幅度有限

。各项实验中拟合理性参数（λ_1，λ_2）的变化趋势进一步支持了这些发现：在结构化、协调的设置中，λ 值较高，体现出类似均衡的行为下接近确定性博弈的特征；而在E4等适应性强、非均衡的情境中，λ 值则较低或存在差异。这些模式凸显了所提方法各自互补的优势：CE-ML适用于结构化协调，LBR-ML则能在不存在相关装置的环境中捕捉有限理性程度的变化以及适应性决策。

VI. 结论

本研究针对 2×2 标准式博弈，求解了不同的逆博弈论问题，以刻画交互行为的不同方面。所提出的 CE-ML 方法在数据符合相关均衡结构时，能够准确且高效地恢复参数；在由 [11] 生成相关均衡或信号介导的情景中，其表现优于现有的 ICE 方法。相比之下，第二种方法 LBR-ML 更能刻画有限理性下的自适应交互，即使协调假设不再成立，仍能保持预测能力。未来工作将把这些框架扩展到多车辆和异质智能体环境，并探索从真实或部分观测到的交通交互中在线估计效用参数。

附录

表 VI 按实验和样本列出的非权重参数 T。CE-ML 对 y_i 相关均衡顶点的混合权重 v_i。

		CE-ML (y_i)					LBR-ML (λ_i)	
Exp.	T	y_1	y_2	y_3	y_4	y_5	λ_1	λ_2
E1	500	0.09	0.10	0.63	0.04	0.14	0.50	0.50
E1	1000	0.14	0.01	0.66	0.02	0.17	2.00	2.00
E1	2000	0.00	0.05	0.52	0.00	0.43	0.50	1.00
E2	500	0.02	0.10	0.00	0.36	0.52	3.00	3.00
E2	1000	0.04	0.10	0.00	0.36	0.51	3.00	3.00
E2	2000	0.00	0.07	0.00	0.36	0.57	3.00	3.00
E3	500	0.00	0.00	0.00	1.00	0.00	–	–
E4	500	0.00	0.00	1.00	0.00	0.00	1.00	3.00

参考文献

[1] R. Chandra、H. Karnan、N. Mehr、P. Stone 和 J. Biswas，“现实世界无结构行人群体中的多智能体逆强化学习”，载于*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）论文集，2025年，即将发表。[2] V. Kuleshov 和 O. Schrijvers，“逆博弈论：在简洁博弈中学习效用”，载于网络与互联网经济学：第*11*届国际会议 *WINE 2015*，阿姆斯特丹（荷兰），*2015*年*12*月*9*日至*12*日，论文集，丛书：计算机科学讲义，E. Markakis 和 G. Sch¨afer 编，第9470卷。Cham：施普林格，2015年，第413–427页。 [在线]。获取地址：https://link.springer.com/chapter/10.1007/978-3-662-48995-6_30

[3]L. Yu、J. Song 和 S. Ermon，“多智能体对抗逆强化学习”，载于第36届国际机器学习会议论文集，丛书：机器学习研究论文集，K. Chaudhuri 和 R. Salakhutdinov 编，第97卷。PMLR，2019年6月9日至15日，第7194–7201页。 [在线]。获取地址：https://proceedings.mlr.press/v97/yu19e.html [4] C. Daskalakis、P. W. Goldberg 和 C. H. Papadimitriou，“计算纳什均衡的复杂性”，《ACM通讯》，第52卷，第2期，第89–97页，2009年。[5] C. F. Camerer，《行为博弈论：策略互动实验》。普林斯顿大学出版社，2003年。[6] Y. Nyarko 和 A. Schotter，“使用引导信念研究信念学习的实验研究”，《计量经济学》，第70卷，1005，2002 3，971 no. pp. – 。[7] R. J. Aumann，“随机化策略中的主观性与相关性”，《数理经济学杂志》，第1卷，第1期，第67–96页，1974年。[8] C. H. Papadimitriou 和 T. Roughgarden，“计算多玩家博弈中的相关均衡”，《ACM杂志》，第55卷，第3期，第14:1–14:29页，2008年。[9] J. Duffy、E. K. Lai 和 W. Lim，“通过相关性实现协调：一项实验研究”，《经济理论》，第64卷，304，2017 2，265no. pp. – 。[10] S. Hart 和 A. Mas-Colell，“导向相关均衡的简单自适应程序”，《计量经济学》，第68卷，1150，2000 5，1127 no. pp. – 。[11] A. Bestick、L. J. Ratliff、P. Yan、R. Bajcsy 和 S. S. Sastry，“用于多玩家非合作环境中效用学习的逆相关均衡框架”，载于第2届ACM高置信度网络系统国际会议论文集，2013年，第9–16页。[12] R. D. McKelvey 和 T. R. Palfrey，“标准式博弈的量化反应均衡”，《博弈与经济行为》，第10卷，第1期，第6–38页，1995年。 [在线]。获取地址：https://www.sciencedirect.com/science/article/pii/S0899825685710238

[13] D. Arbis 和 V. V. Dixit，“纳入冲突风险的变道博弈论模型”，《事故分析与预防》，第125卷，第158–164页，2019年。[14] Y. Zhang 和 J. D. Fricker，“将冲突风险纳入行人与机动车驾驶员互动：一种博弈论方法”，《事故分析与预防》，第159卷，文章编号106254，2021年。[15] M. A. Costa-Gomes、V. P. Crawford 和 B. Broseta，“标准式博弈中的认知与行为：一项实验研究”，《计量经济学》，第69卷，第5期，第1193–1235页，2001年。[16] Y. S. Razin 和 K. M. Feigh，“承诺相互依赖：博弈论对人机信任的启示”，《Paladyn：行为机器人学杂志》，第12卷，第1期，第481–502页，2021年。

[17] [ ] 10 1515 2021 0031 在线。获取地址：https://doi.org/ . /pjbr- - A. Calvo-Armengol，“博弈的相关均衡集合”，巴塞罗那经济学工作论文系列，巴塞罗那，技术报告79，2003年10月。 [在线]。获取地址：https://bse.eu/sites/ default/files/working paper pdfs/79.pdf [18] E. Blume，L. “策略互动的统计力学”，《博弈与经济行为》，第5卷，第3期，第387–424页，1993年。[19] P. A. Lopez、M. Behrisch、L. Bieker-Walz、J. Erdmann、Y.-P. Fl¨otter¨od、R. Hilbrich、L. L¨ucken、J. Rummel、P. Wagner 和 E. Wießner，“使用SUMO进行微观交通仿真”，载于第21届IEEE智能交通系统国际会议论文集。IEEE，2018年。 [在线]。获取地址：https://elib.dlr.de/124092/ [20] T. Tatarenko，《无记忆多智能体系统中的博弈论学习与分布式优化》。瑞士Cham：施普林格国际出版，2017年。[21] E. Seneta，《非负矩阵与马尔可夫链》，修订版。纽约：施普林格，2006年。[22] L. E. Ortiz、R. E. Schapire 和 S. M. Kakade，“最大熵相关均衡”，载于《人工智能与统计学》。PMLR，2007年，第347–354页。[23] A. H. Kalantari、Y. Yang、Y. M. Lee、N. Merat 和 G. Markkula，“无信号交叉口的驾驶员与行人互动并不符合纳什均衡”，《IEEE Access》，第11卷，110 723，2023 110 707pp. – 。[24] M. Liu、Y. Wan、F. L. Lewis、S. Nageshrao 和 D. Filev，“自动驾驶车辆的三级博弈论决策框架”，《IEEE智能交通系统汇刊》，第23卷，第11期，第20 298–20 308页，2022年。