自动驾驶按需出行系统中用于联合定价与车队再平衡的竞争性多运营商强化学习

Emil Kragh Toft[1], Carolin Schmidt[2],Daniele Gammelli[3],Filipe Rodrigues[1]

arXiv:2603.05000v1 [cs.LG] 5 Mar 2026

摘要——**自动驾驶按需出行系统有望通过提供价格实惠的按需服务，满足不断增长的出行需求，从而革新城市交通。然而，现实中的自动驾驶按需出行市场具有竞争性，多个运营商将通过策略性定价和车队部署争夺乘客。尽管强化学习在优化单运营商自动驾驶按需出行控制方面展现出潜力，现有研究却未能反映竞争性市场动态。我们通过提出一个多运营商强化学习框架，研究竞争对策略学习的影响：两个运营商同时学习定价与车队再平衡策略。通过整合离散选择理论，我们使乘客分配和需求竞争能够从效用最大化决策中内生地产生。基于多个城市真实世界数据开展的实验表明，竞争会从根本上改变学习到的行为，与垄断环境相比，导致价格更低、车队布置模式也有所不同。值得注意的是，我们证明，基于学习的方法能够适应竞争带来的额外随机性；竞争智能体在考虑到竞争对手策略部分不可观测的情况下，仍能成功收敛至有效策略。**

一、引言

城市出行系统正面临快速城市化和消费者预期不断变化带来的日益增长的压力。私人汽车拥有量与僵化的公共交通之间的差距，推动了网约车服务的迅速发展，从根本上改变了城市出行 [1]。自动驾驶按需出行系统（AMoD）有望进一步革新交通运输，通过大规模提供经济实惠的门到门服务 [2]。然而，与受监管的公共交通垄断不同，自动驾驶按需出行市场很可能具有竞争性，多家运营商将部署车队并争夺乘客。

在此类市场中，运营商必须在预判竞争对手行动的同时，策略性地制定价格并部署车队；例如，一家运营商降价会使其他运营商流失乘客。这些互动构成了复杂的竞争环境，最优策略取决于竞争对手的行为。

强化学习（RL）已成为自动驾驶按需出行控制的一种强大方法，能够从数据中推导出可扩展且接近最优的策略，而无需显式建立需求或竞争对手策略模型。然而，现有强化学习研究主要关注集中式的单运营商场景，研究车辆再平衡 [3], [4], 、定价 [5], [6], 或两者 [7],；而现实中的自动驾驶按需出行系统将在竞争环境中部署。

[1]丹麦技术大学，丹麦s233791@student.dtu.dk, rodr@dtu.dk[2]慕尼黑工业大学，德国carolin.schmidt@tum.de[3]斯坦福大学，美国gammelli@stanford.edu

本研究通过构建一个用于联合定价与车队再平衡的竞争性多运营商强化学习框架，探究竞争对策略学习的影响。我们对两家运营商进行建模，研究它们如何在争夺乘客需求的同时学习并调整策略；乘客需求通过一个对价格、出行时间和工资敏感的选择模型进行分配。主要贡献如下：

- 我们将双运营商自动驾驶出行服务的竞争问题建模为控制问题，两家独立运营商联合学习定价和再平衡策略，从而将联合控制强化学习从垄断市场拓展到竞争市场。
- 我们将一种考虑工资依赖型价格敏感度的乘客选择机制整合到学习循环中，使运营商之间的需求分配能够根据各自的行动内生地产生。
- 利用来自多个城市的真实出租车行程数据，我们通过实证证明竞争环境下的同时学习能够收敛，并分析竞争相较于垄断控制如何改变学习到的策略、服务质量和市场效率。

II. 文献综述

自动驾驶按需出行系统的早期研究依赖排队论模型 [8] 和用于车队再平衡的模型预测控制 [9], ，随后又通过引入鲁棒优化来处理需求不确定性 [10]。尽管这些方法具有理论基础，但当需求偏离假设分布或计算预算不足时，它们便难以奏效 [11]。近年来，强化学习已成为一种数据驱动的替代方法，无需显式的需求或交通模型，便可通过交互直接学习策略 [4]。Gammelli 等人 [3] 提出了一种用于再平衡的分层框架，结合传统优化与强化学习的优势，并利用图神经网络挖掘交通网络的空间结构。借助这些工具，强化学习能够高效扩展至大规模的真实城市出行场景 [12]。

尽管这些方法侧重于供给，动态定价可通过塑造需求来补充再平衡。早期的定价研究采用了基于均衡的方法 [13] 和运筹学公式 [14],，而双层模型 [15] 则联合优化运营商和乘客的决策。然而，将定价或再平衡孤立处理，无法发挥二者的协同效应 [7]。特别是，近期一项基于强化学习的联合再平衡与定价框架表明，联合控制始终优于单独策略 [7]。

多运营商竞争已通过博弈论定价模型 [16], 基于智能体的仿真 [17],以及服务质量竞争框架 [18]展开研究。然而，

这些研究依赖解析均衡或静态模型，并且将定价与车队管理割裂开来，分别进行研究。尽管强化学习已在单运营商自动驾驶出行服务场景中得到探索，本研究关注竞争性的多运营商场景，旨在联合学习竞争性再平衡与定价策略，并纳入内生的价格响应型需求。具体而言，本研究构建了一个双运营商强化学习框架，让相互竞争的运营商在争夺需求的同时同步学习策略，从而填补这一研究空白，并探讨竞争对基于强化学习的策略学习的影响。

III. 设计与实现

我们将自动驾驶出行服务环境表示为有向图$\mathcal{G} = (\mathcal{V}, \mathcal{E})$，其中 $\mathcal{V}$ 包含 N_v 个空间区域，每个区域均以指定的上车／下车站点为中心。总车队由两家独立运营商分配并控制，分别拥有M_0 辆和 M_1 辆自动驾驶车辆。为便于说明，我们重点讨论双运营商的情形，但该框架可轻松扩展至更多运营商。两家运营商在同步的离散时间范围 $\mathcal{T} = \{1, \ldots, T\}$ 内运行，时间步长为3分钟。从区域 i到区域 j 的移动沿最短路径进行，需要 $\tau_{ij} \in \mathbb{Z}_+$ 个时间步，并产生随时间变化的运营成本 c^t_{ij}。

在时间 t ，站点 i的可用供给由各车队的空闲车辆数量$m^t_{i,0}$ 和$m^t_{i,1}$ 表示。在需求侧，到达起点 i 的乘客通过离散选择模型评估两家自动驾驶出行服务运营商及一种替代交通方式：根据行程价格、乘客收入和预计旅行时间计算各选项的效用 U ，将其映射为选择概率，并从类别分布中抽样。乘客选择某家运营商后，便进入先到先服务（FCFS）队列。等待时限设为6分钟；若在此时间内没有车辆派出，乘客便离开系统。

基于这一结构，以下各节将对运营商之间的互动进行形式化描述。我们首先将定价与再平衡任务建模为马尔可夫决策过程（MDP），然后详细介绍一种旨在捕捉双运营商学习中固有空间依赖关系的图神经网络架构。

A. 竞争性自动驾驶出行服务控制

我们研究一种双运营商自动驾驶出行服务（AMoD）环境，其中两家独立运营商在共享环境中分别寻求最大化自身利润。控制架构遵循 [3], [7], 的分层形式，从而能够扩展到现实世界的城市场景 [12]。我们将该框架调整到竞争环境中，以研究竞争对基于强化学习的策略学习的影响。在实践中，控制架构包含三个步骤，具体如附录所示。具体而言，在每个时间步 t，仿真按以下三个步骤进行：**(1) 定价与再平衡：** 每个运营商的策略生成基于出发地的价格系数和期望空闲车辆分布，然后将价格标量与根据历史数据估计的起讫点基础价格相乘，转换为票价。**(2) 需求分配：** 随机选择模型生成乘客请求，并通过价格依赖型选择模型将其分配给各运营商分配到的乘客在其出发区域排队，并按照先到先服务顺序等待服务，同时受系统级最大等待时间限制。**(3) 再平衡执行：** 通过求解最小费用流问题，将每个运营商的期望分布映射为再平衡行动（见附录）。各运营商分别独立执行流量，更新车辆状态和队列，系统转移至 $t{+}1$，并返回新的状态和奖励。

B. 自动驾驶出行控制问题作为马尔可夫决策过程

我们将多运营商自动驾驶出行服务控制建模为一个马尔可夫决策过程，其组成部分（$\mathcal{S}$，$\mathcal{A}$，$\mathcal{P}$，$\mathcal{R}$）定义如下。

状态空间 ($\mathcal{S}$)：运营商 o 在时间步 $t \in \mathcal{T}$的状态$\mathbf{s}_{t,o}$ 包括：网络邻接矩阵$\mathbf{A}$，各区域的自身闲置车辆数量 i，即 $m^t_{i,o} \in [0, M_o]$；规划时域内在途车辆 H_p, $\{m^{t'}_{i,o}\}_{t'=\tau,\ldots,\tau+H_p}$；上一个时间步的自身价格和竞争对手价格 $p^{t-1}_{i,o}$, $p^{t-1}_{i,o'}$，适用于所有 i，$j \in \mathcal{V}, o' \neq o$；自身队列长度 $q^t_{i,o}$，以及上一个时间步的自身需求 $d^{t-1}_{i,o}$，适用于所有 $i \in \mathcal{V}$。运营商之间不共享需求或车辆位置数据，但可以观测竞争对手的价格。

动作空间($\mathcal{A}$)：每个运营商 o输出基于出发地的价格系数 $\rho^t_{i,o} \in (0, 1]$ 以及期望的闲置车辆分布 $w^t_{i,o} \in [0, 1]$ 其中 $\sum_{i=1}^{N_v} w^t_{i,o} = 1$，后者用于求解最小费用再平衡问题。选择基于出发地的价格系数是根据先前研究 [7], ，该研究表明，与基于起讫点票价的价格系数相比，这种方法收敛更快且性能相当。联合动作为 $\mathbf{a}_{t,o} = [\mathbf{w}_{t,o}, \rho_{t,o}]$ 其中 $\mathbf{w}_{t,o} = [w^t_{1,o}, \ldots, w^t_{N_v,o}]$ 且 $\rho_{t,o} = [\rho^t_{1,o}, \ldots, \rho^t_{N_v,o}]$。随后按如下方式计算起讫点票价：

$$p^t_{i,j,o} = \beta \cdot \rho^t_{i,o} \cdot \overline{p}^t_{i,j}, \tag{1}$$

其中 $p^t_{i,\neg j}$ 是历史参考价格（根据数据得出），而 β 是价格水平的上限（我们在实验中设定 $\beta = 2$，得到 $p^t_{i,j,o} \in (0, 2\overline{p}^t_{i,j}]$）。

奖励 ($\mathcal{R}$)：每个运营商都最大化自身利润，利润定义为行程收入减去运营成本：

$$r_o = \sum_{i,j \in \mathcal{V}} x^t_{i,j,o}\left(p^t_{i,j,o} - c^t_{i,j,o}\right) - \sum_{(i,j)\in\mathcal{E}} y^t_{i,j,o}\, c^t_{i,j,o}, \tag{2}$$

其中，$x^t_{i,j,o}$ 表示已服务的乘客，$y^t_{i,j,o}$ 表示重新调配的车辆，$c^t_{i,j,o}$ 表示从 i到 j的单位车辆行驶成本。

动态过程 ($\mathcal{P}$)：状态转移由随机需求、队列更新和车辆可用量守恒驱动。邻接矩阵 $\mathbf{A}$ 是外生且静态的。

需求生成与模式选择：每位乘客 k 的选择通过多项Logit模型，根据效用在运营商 0、运营商 1 或替代选项（$\emptyset$）之间进行抽样

$$U^t_{k,i,j,o} = \beta_0 - \beta_t \cdot v_k \cdot \tau^t_{i,j} - \frac{\bar{v}}{v_k}\, p^t_{i,j,o}, \tag{3}$$

其中 β_0 为基准偏好，β_t 为旅行时间的边际负效用（设为0.71，反映相对于货币成本对时间低估29%，与 [19]一致），v_k为乘客的小时工资，$\bar{v}$ 为全场景平均工资。比率 $\bar{v}/v_k$ 刻画收入效应。备选方案的效用归一化为零。对于每个起讫点对 (i, j)

以及时间 t，参考需求$\bar{d}^t_{i,j}$ 和价格$\bar{p}^t_{i,j}$ 均从历史数据 [20], [21]中抽样。潜在需求池缩放至$2\bar{d}^t_{i,j}$，并对 β_0 进行校准，使历史票价下运营商的总需求等于参考需求。选择模型对价格标量的敏感性见附录。工资数据来源于美国人口普查局收入统计 [22]，，其中旧金山场景的工资使用2008年至2011年的全国美国通胀率进行调整，以便由2011年数据得到2008年等价估计。

队列与车辆动态过程：运营商 o 在区域 i 的队列演变如下

$$q^t_{i,o} = q^{t-1}_{i,o} + \sum_{j\in\mathcal{V}\setminus\{i\}} \left(d^t_{i,j,o} - x^t_{i,j,o}\right), \tag{4}$$

通过平衡新增到达量与成功匹配量来更新。由于自动驾驶车辆会完全遵从调度指令，空闲车辆数量更新如下

$$m^t_{i,o} = m^{t-1}_{i,o} + \sum_{j\in\mathcal{V}\setminus\{i\}} v^{\text{arr},t}_{j,i,o} - \sum_{j\in\mathcal{V}\setminus\{i\}} \left(x^{t-1}_{i,j,o} + y^{t-1}_{i,j,o}\right), \tag{5}$$

其中，$v^{\text{arr},t}_{j,i,o}$ 表示在时间 t 从 j 到达 i 的车辆数量（完成乘客行程或在先前时段执行再平衡调度），流出量则包括匹配后出发的车辆 $x^{t-1}_{i,j,o}$ 以及再平衡调度的车辆 $y^{t-1}_{i,j,o}$。

C. 模型架构

总体而言，每个运营商的目标是学习一种政策 $\pi(a_t|s_t)$，通过与马尔可夫决策过程的交互来最大化其期望累积奖励。在我们的设置中，每个运营商 o 都维护一个独立的行动者 $\pi_{\theta_o}(\cdot \mid \mathbf{s}_t)$ 和评论家 $V_{\phi_o}(\mathbf{s}_t)$，运营商之间不共享参数。网络以状态 $\mathbf{s}_t$ （第 III-B 节）为输入，采用卷积神经网络（GCN）编码 [3]，[7] ，随后接入全连接层（见附录）。评论家通过全局求和汇总节点嵌入，以生成标量值 $V_{\phi_o}(\mathbf{s}_t)$。行动者通过 Softplus 输出严格为正的参数，用于定义随机策略。对于联合定价与再平衡，每个区域 i 都产生参数 α^t_i 和 β^t_i，用于贝塔分布，并从中采样基于起点的价格标量：$p^t_{i,o} \sim$ 贝塔$(\alpha_i,\ \beta_i)$；另有一个浓度标量 γ^t_i，它们共同构成 $\boldsymbol{\gamma}^t \in \mathbb{R}^{N_v}$，用于参数化狄利克雷分布，从中采样期望的空闲车辆份额：$\mathbf{w}^t_o \sim$ 狄利克雷$(\boldsymbol{\gamma}^t)$。对于仅定价或仅再平衡控制，行动者分别只输出贝塔分布或狄利克雷分布的参数。

IV. 实验

本节从两个维度对所提出的框架进行实证评估：(i) 比较旧金山、华盛顿特区和纽约市曼哈顿南部的性能，针对仅再平衡、仅定价和联合控制模式，对比垄断式单运营商控制与竞争式双运营商场景；以及 (ii) 对车队规模、非对称车队分布和区域工资异质性进行敏感性分析，并考察它们对均衡结果的影响。

所有实验均使用从 [20], 改编而来的离散时间仿真环境，涵盖晚间高峰时段（19:00–20:00）。需求采用泊松分布建模，其率参数根据现实世界的出租车行程数据[20], [21]推导得到。运营商使用优势行动者评论家（A2C）进行训练，采用第三-C节所述的基于图卷积网络（GCN）的架构，训练至收敛后通过10次测试运行进行评估。我们报告平均值，并在括号中给出标准差。 [1]

A. 多城市性能分析

我们在三个具有不同特征的城市环境中评估性能，特征汇总于表 I。空间需求变异性由变异系数（CV）衡量，定义为区域需求的标准差与均值之比。我们进一步报告两个非学习基准差距：无控制（“NC”）基线，对应于在无再平衡或定价干预情况下的系统性能；以及均匀分布启发式（“UD”），其目标是在固定价格下将闲置车辆均匀分布到整个网络。奖励差距和服务差距衡量均匀分布相对于无控制在奖励和已服务乘客方面的相对改进，作为再平衡所获改进的近似。

表 I：用于比较分析的场景特征。

City	每小时 Wage	节点	Veh.	需求	奖励 Gap	已服务 Gap	CV
San 旧金山	$17.76	10	374	5490	46.72%	63.04%	1.31
华盛顿 DC	$25.26	18	1096	16881	15.14%	48.13%	1.26
纽约曼哈顿南部	$22.77	12	650	21270	10.03%	20.83%	0.69

注：车辆：车辆数；CV：需求变异系数。

1) 单运营商垄断性能：表 II 和表 III 展示了单运营商场景下不同控制策略的总奖励。除无控制和均匀分布启发式方法外，基于学习的方法还包括仅再平衡、仅定价，以及同时控制这两项决策的联合策略。在三个城市中，联合策略的表现均优于所有基线方法和单一模式策略。在旧金山，联合控制的表现比单独再平衡提高了 22.4%，比单独定价提高了 42.4%。在需求变异性较低的城市中，这种提升幅度会减小：相较于再平衡，华盛顿特区的收益降至 5.3%，而纽约市曼哈顿南部的收益则微乎其微（ 0.02% ）。在华盛顿特区和纽约市曼哈顿南部，无论采用仅定价策略还是联合控制策略，动态定价都会导致垄断性涨价，与采用固定价格的再平衡相比，已服务乘客人数因此减少。

2）双运营商竞争性能：表 IV 展示了双运营商场景下的总奖励。与单运营商情况不同，没有一种控制模式能在所有城市中占据优势。在旧金山，联合控制取得

[1]补充材料、代码和超参数见 https:// github.com/emto17ab/dual_agent_rl

表 II：单运营商（垄断）利润 [$]。最优策略以粗体显示。我们报告 10 次测试运行的平均值，括号内为标准差

City	NC	UD	Reb.	定价	联合
旧金山	6345.49 (251.60)	9470.06 (336.71)	10037.29 (337.03)	8629.30 (193.96)	**12286.03 (166.74)**
华盛顿特区	13153.97 (324.58)	15630.03 (388.50)	15976.21 (395.37)	13972.47 (301.58)	**16816.04 (287.07)**
纽约市曼哈顿南部	16283.02 (487.67)	18221.62 (260.07)	18499.21 (291.51)	17004.39 (504.23)	**18503.52 (324.97)**

了最高奖励（10,415$），而华盛顿特区则更适合再平衡（15,620$）。值得注意的是，仅再平衡策略可作为固定价格基准，但在实际中并不现实，因为两家运营商采用相同价格，彼此之间不存在价格竞争。在纽约市曼哈顿南部，仅定价控制取得最高奖励（18,632$），超过了另外两种模式。这种差异表明，在高变异环境中，车队部署是主要的竞争手段；而在稳定的高密度场景中，定价可以有效发挥隐式再平衡机制的作用。此外，联合策略在纽约市南部表现较差，凸显了竞争环境下联合控制问题的复杂性更高：作为一种长期决策，再平衡需要考虑未来需求的不确定性，而未来需求会受到竞争对手未被观察到的策略的显著影响。

表 IV：双运营商联合利润 [$]。最优策略以粗体显示。

City	NC	UD	Reb.	定价	联合
旧金山	6391.2 (232.4)	9352.6 (258.9)	10065.9 (248.7)	8919.0 (160.4)	**10415.1 (178.8)**
华盛顿特区	13172.3 (344.6)	15113.5 (325.8)	**15619.5 (325.7)**	12820.8 (318.2)	13460.6 (247.6)
纽约市曼哈顿南部	16048.4 (416.6)	17652.0 (277.5)	18133.7 (330.8)	**18631.6 (376.7)**	17481.1 (258.7)

表 V 提供了双运营商场景的详细指标。几个值得注意的模式如下。首先，与垄断情形相比，竞争会压低价格。其次，双运营商场景下的等待时间通常会增加，这反映了车队管理分散所带来的低效。两家运营商之间的利润分配大致均衡，典型差异不到 2%。

3）垄断与竞争比较：比较不同市场结构下表现最佳的策略，可以看出竞争所导致的利润损失。在旧金山，垄断联合控制实现12,286美元，而最佳双运营商配置实现10,415美元，降幅为15.2%。华盛顿特区的降幅为7.1%。在纽约市曼哈顿南部，双运营商定价策略实现的利润略高于垄断联合控制。这可能反映出竞争性定价在这一高密度环境中刺激了额外需求，不过该差异（0.7%）可忽略不计。总体而言，与垄断情形相比，定价竞争压低了价格，且在所有情景中，旧金山最高达27%。竞争所导致的利润损失幅度似乎与需求变异性相关。高变异系数环境在竞争下损失更大，因为碎片化车队管理加剧了将供给与波动需求相匹配的难度。在低变异系数的纽约市场景中，主要竞争维度从再平衡效率转向基于价格的市场份额竞争，此时碎片化的代价更小。

B. 纽约市曼哈顿南部的竞争性策略分析

表 IV 的结果并未直接揭示任何特定的定价行为，因为两家运营商的平均价格都没有持续大幅低于对方。然而，对纽约市南部仅定价策略进行细致分析后发现，运营商已经学会在特定区域略微策略性地压低价格，以获取需求，同时避免因过度压价而造成利润的无效损失。如图 6 所示。在 $t = 0$，操作员 1 选择性地在四个高需求区域压低价格，而操作员 0 则在其余区域压低价格。这些策略还体现出显著的区域差异：北部高需求区域与南部低需求区域的价格标量差异最高可达 83%。关于再平衡策略，图 2 显示，两家运营商的再平衡模式都紧密跟随需求模式；操作员 0 的再平衡更有针对性，而操作员 1 则选择了更均衡的策略。



图 1：纽约市曼哈顿南部仅定价策略在时间步 0 的初始定价系数。

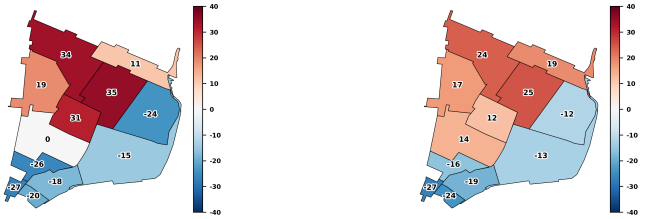


图 2：联合控制策略的净再平衡流，展示所有时间步内车辆移动的累计情况。红色 =表示净接收方，蓝色 = 表示净发送方。

C. 定价信息的影响

我们比较运营商可以观察和无法观察竞争对手价格的情景。我们的实验表明，系统

表 III：单运营商（垄断）表现。“等待时间”指乘客等待时间。最佳结果以粗体显示。

	旧金山			华盛顿特区			纽约市曼哈顿南部		
政策	Reb.	定价	联合	Reb.	定价	联合	Reb.	定价	联合
奖励 [$]	10037.29 (337.03)	8629.30 (193.96)	**12286.03 (166.74)**	15976.21 (395.37)	13972.47 (301.58)	**16816.04 (287.07)**	18499.21 (291.51)	17004.39 (504.23)	**18503.52 (324.97)**
再平衡成本 [$]	877.08 (25.30)	—	650.04 (17.36)	3475.65 (152.04)	—	2975.85 (83.57)	1436.70 (85.48)	—	1712.25 (101.98)
再平衡行程	523.70 (12.59)	—	384.00 (13.70)	1085.80 (34.25)	—	955.50 (25.45)	295.90 (20.40)	—	361.10 (23.56)
价格标量	—	0.75 (0.00)	0.88 (0.00)	—	1.20 (0.00)	1.02 (0.00)	—	1.07 (0.00)	1.02 (0.00)
等待时间 [分钟]	0.85 (0.14)	1.49 (0.08)	1.57 (0.15)	1.03 (0.09)	1.25 (0.15)	0.64 (0.09)	1.91 (0.06)	1.48 (0.08)	1.71 (0.10)
队列	2.03 (0.34)	4.44 (0.26)	5.40 (0.41)	5.56 (0.44)	4.51 (0.42)	3.03 (0.39)	14.20 (0.48)	9.09 (0.40)	12.07 (0.65)
已服务需求	960.90 (30.48)	816.60 (15.14)	1273.70 (16.36)	4182.40 (59.60)	2309.50 (49.93)	3860.30 (44.29)	3571.10 (36.75)	2777.80 (78.12)	3528.10 (39.21)
总需求	1092.10 (42.86)	1298.90 (37.03)	1685.60 (35.72)	4753.00 (76.79)	2975.40 (43.52)	4129.40 (54.38)	4705.70 (73.36)	3590.10 (73.90)	4420.00 (77.77)

表 V：双运营商性能指标。表现最佳的策略以粗体显示。

	旧金山			华盛顿特区			纽约市曼哈顿南部		
政策	Reb.	定价	联合	Reb.	定价	联合	Reb.	定价	联合
总奖励 [$]	10065.9 (248.7)	8919.0 (160.4)	**10415.1 (178.8)**	**15619.5 (325.7)**	12820.8 (318.2)	13460.6 (247.6)	18133.7 (330.8)	**18631.6 (376.7)**	17481.1 (258.7)
奖励算子 0	5071.5 (178.5)	4191.5 (77.1)	5190.7 (116.7)	7795.7 (233.7)	6386.6 (189.6)	6726.0 (135.8)	9010.8 (242.6)	9351.6 (161.4)	8717.0 (249.1)
奖励算子 1	4994.4 (159.8)	4727.6 (149.5)	5224.4 (126.3)	7823.8 (117.2)	6434.2 (255.9)	6734.6 (171.4)	9123.0（341.4）	9280.1（265.1）	8764.1（223.5）
总报销费用 [$]	1006.8（24.0）	—	549.1（20.7）	4120.8 (120.2)	—	3368.1 (125.2)	1600.2 (95.4)	—	1379.0 (83.7)
返利 成本 操作员 0	498.5 (19.3)	—	296.7 (10.6)	2094.8 (88.5)	—	1682.7 (72.1)	813.5 (73.6)	—	699.6 (76.7)
返程成本操作员1	508.3 (19.0)	—	252.4 (13.3)	2026.0 (59.2)	—	1685.4 (83.3)	786.8 (107.0)	—	679.4 (63.1)
返程总次数	593.8 (14.4)	—	323.8 (13.5)	1221.9 (47.4)	—	911.1 (37.2)	338.4 (21.2)	—	284.7 (19.0)
返利 行程 操作员 0	294.0 (11.0)	—	172.3 (5.3)	618.2 (37.0)	—	457.2 (20.4)	173.5 (18.2)	—	144.9 (17.8)
篮板球行程操作员 1	299.8 (9.7)	—	151.5 (9.8)	603.7 (15.3)	—	453.9 (21.6)	164.9 (24.2)	—	139.8 (14.3)
总服务需求	972.3 (23.4)	865.8 (14.4)	1378.9 (17.1)	4239.5 (46.8)	2892.8 (67.6)	4414.4 (33.0)	3535.8 (39.4)	3337.6 (52.3)	3561.1 (29.0)
已服务需求	488.4 (16.1)	403.6 (6.7)	694.5 (12.3)	2124.2 (34.5)	1447.5 (43.7)	2205.9 (18.5)	1758.1 (29.4)	1674.3 (19.1)	1773.9 (26.1)
已服务需求	483.9 (15.1)	462.2 (13.8)	684.4 (11.7)	2115.3 (16.7)	1445.3 (56.6)	2208.5 (21.4)	1777.7 (39.5)	1663.3 (38.7)	1787.2 (24.0)
价格标量算子 0	—	0.72 (0.00)	0.64 (0.00)	—	0.97 (0.00)	0.90 (0.00)	—	0.95 (0.00)	0.96 (0.00)
价格标量算子 1	—	0.72 (0.00)	0.67 (0.00)	—	0.98 (0.00)	0.90 (0.00)	—	0.95 (0.01)	0.96 (0.00)
等待时间 运营商 0 [分钟]	0.84 (0.18)	1.57 (0.07)	1.95 (0.11)	0.88 (0.06)	1.73 (0.07)	1.73 (0.17)	1.80 (0.08)	1.51 (0.08)	2.03 (0.10)
等待时间 运营商 1 [分钟]	0.86 (0.17)	1.68 (0.12)	1.85 (0.11)	0.87 (0.12)	1.77 (0.06)	1.71 (0.12)	1.78 (0.07)	1.45 (0.08)	2.06 (0.08)
队列算子 0	2.24 (0.65)	3.44 (0.84)	5.29 (0.82)	1.98 (0.85)	6.11 (0.57)	5.84 (0.89)	8.41 (1.12)	5.69 (0.73)	10.4 (1.3)
队列操作员 1	2.71 (0.63)	3.87 (0.52)	3.23 (0.66)	2.04 (0.73)	5.89 (0.83)	5.53 (1.25)	7.75 (1.05)	4.96 (0.67)	10.8 (1.2)
总需求	1112.5 (43.3)	1519.5 (34.2)	2154.1 (40.9)	4746.9 (76.8)	4990.6 (73.0)	5682.0 (92.1)	4683.0 (73.9)	4311.7 (65.1)	5231.3 (61.6)
运营商 0 的需求	551.9 (20.2)	721.3 (25.2)	1192.4 (32.5)	2379.4 (54.5)	2544.8 (49.7)	2859.9 (51.2)	2336.2 (56.0)	2145.4 (36.7)	2581.1 (59.5)
运营商 1 的需求	560.6 (30.9)	798.2 (14.5)	961.7 (21.3)	2367.5 (36.6)	2445.8 (29.0)	2822.1 (44.0)	2346.8 (39.4)	2166.3 (34.7)	2650.2 (42.1)

在所有策略模式下，性能对这一信息条件都具有稳健性，收敛奖励几乎相同。这表明，在从头开始同时学习的情景中，竞争对手价格可能带来的是噪声，而非可操作的信号。完整训练曲线见附录。

D. 车队规模敏感性分析

我们比较了纽约市曼哈顿南部在不同车队规模（由各运营商平均分配）下，联合策略、无控制和均匀分布基线的系统性能。随着车队规模增大，联合策略会动态降低价格标量（从 1.04 降至 0.71），以维持更大车队的利用率，服务的乘客数量最多比 NC 高 64%。总奖励在 1，050 辆车时达到峰值，之后随着再平衡成本超过收入增长而下降，这表明车队扩张的利润正在递减。完整结果见附录。

E. 车队划分敏感性分析

我们在纽约市曼哈顿南部评估联合策略，考察两家运营商的五种不对称车队划分，比例从 5:5 到 1:9 不等。总奖励在 4:6 的划分下最高（17,567 美元），在不对称程度更高时则有所下降；而总服务需求在不同配置下保持稳定（3,500–3,577）。规模较小的运营商通过提高价格来弥补运力有限（从 5:5 时的 0.96 升至 1:9 时的 1.10），而另一家运营商凭借规模更大的车队，能够持续以更低价格竞争（0.93–0.96）。这些结果表明，车队规模失衡会减弱直接的价格竞争。完整结果见附录。

F. 区域工资异质性

最后，我们通过在纽约市曼哈顿南部开展实验，评估乘客收入的空间差异所带来的影响。该地区各区域的工资从 10 美元到 30 美元不等，两家运营商平分总计 650 辆的车队。值得注意的是，再平衡流体现了空间适应性：两家运营商都将车辆从低收入地区调往需求集中的高收入地区，同时提高定价系数，以利用更高的支付意愿。完整结果见附录。

V. 总结与结论

本文提出了一个面向自动驾驶按需出行系统中联合定价与车队再平衡的竞争性多运营商强化学习框架，以研究竞争如何影响运营商策略学习。通过在学习循环中嵌入离散选择模型，乘客分配通过效用最大化内生形成，使定价行动与需求竞争相互关联。至关重要的是，我们通过实证表明，基于学习的方法对于竞争带来的额外随机性仍具有鲁棒性。竞争智能体会收敛到稳定的奖励，并学习有效的策略，在考虑未观测到的竞争者行为的同时对车队进行再平衡并设定动态价格。

在三个城市开展的实验得出了若干重要发现。在垄断情形下，联合定价与再平衡始终能够最大化利润。竞争从根本上改变了占优策略：联合控制在垄断情形下表现出色，而在竞争市场中，专门化策略可能更具优势。

竞争会压低价格，使乘客受益于较低的票价，但多运营商车队管理会延长等待时间，体现出复杂的福利效应。敏感性分析表明，运营商会根据车队规模变化动态调整；不对称车队划分会减弱直接价格竞争，小型运营商提高价格，而大型运营商则压低价格；工资异质性则会推动车队重新部署至高收入地区，并提高价格以利用更高的支付意愿。

这些发现为未来工作指出了若干方向：将超过最大阈值的细粒度等待时间纳入乘客效用，从而引入盈利能力与服务质量之间的直接权衡；研究由非对称架构或学习参数引发的涌现行为；以及显式建模合谋行为，以分析其对定价和消费者福利的影响。

参考文献

[1] A. A. Ceder，“城市出行与公共交通：未来展望与综述”，《国际城市科学杂志》，第25卷，第4期，第455–479页，2021年。[2] Z. Wang与S. Li，“自动驾驶与传统网约车平台之间的竞争：市场均衡与技术转移”，《交通研究*C*辑：新兴技术》，第165卷，第104728页，2024年。[3] D. Gammelli、K. Yang、J. Harrison、F. Rodrigues、F. C. Pereira与M. Pavone，“面向自动驾驶按需出行系统的图神经网络强化学习”，载于*IEEE*决策与控制会议（*CDC*），2021年，第2996–3003页。

[4] J. Wen、J. Zhao与P. Jaillet，“共享按需出行系统的再平衡：一种强化学习方法”，载于*IEEE*智能交通系统国际会议（*ITSC*），2017年，第220–225页。[5] Z. Lei与S. V. Ukkusuri，“网约车系统动态定价的可扩展强化学习方法”，《交通研究*B*辑：方法论》，第178卷，第102848页，2023年。[6] C. Chen、F. Yao、D. Mo、J. Zhu与X. M. Chen，“基于强化学习的网约车平台时空定价”，《交通研究*C*辑：新兴技术》，第130卷，第103272页，2021年。[7] X. Li、C. Schmidt、D. Gammelli与F. Rodrigues，“学习自动驾驶按需出行的联合再平衡与动态定价策略”，《*IEEE*智能交通系统汇刊》，第26卷，第10期，第16619–16634页，2025年。[8] R. Zhang与M. Pavone，“机器人按需出行系统的控制：排队理论视角”，《国际机器人研究杂志》，第35卷，第1-3期，第186–203页，2016年。[9] G. C. Calafiore、C. Bongiorno与A. Rizzo，“按需出行系统再平衡的鲁棒模型预测控制方法”，《控制工程实践》，第90卷，第169–181页，2019年。[10] X. Guo、N. S. Caros与J. Zhao，“不确定需求下网约车系统中的鲁棒匹配集成车辆再平衡”，《交通研究*B*辑：方法论》，第150卷，第161–189页，2021年。[11] R. Iglesias、F. Rossi、K. Wang、D. Hallac、J. Leskovec与M. Pavone，“自动驾驶按需出行系统的数据驱动模型预测控制”，载于*IEEE*机器人与自动化国际会议（*ICRA*），2018年，第6019–6025页。[12] L. Tresca、C. Schmidt、J. Harrison、F. Rodrigues、G. Zardini、D. Gammelli与M. Pavone，“通过强化学习实现大规模机器人出租车车队协调”，2025年，arXiv:2504.06125。[13] G. P. Cachon、K. M. Daniels与R. Lobel，“动态加价在具有自主排班运力的服务平台中的作用”，《运营战略电子期刊》，2016年。[14] Y. Guan，“共享按需出行的设计与优化：动态路径规划与动态定价”，博士学位论文，麻省理工学院，2021年。[15] W. Tang、H. Wang、Y. Wang、C. Chen与X. Chen，“网约车平台空间定价策略的双层优化模型”，《先进交通杂志》，第2022卷，第1–22页，2022年。[16] Y. Yang与M. Ramezani，“双寡头网约车市场中的日内竞争”，载于第*12*届交通分析三年研讨会，2025年。[17] R. Engelhardt、P. Malcolm、F. Dandl与K. Bogenberger，“自动驾驶拼车服务的竞争与合作：经纪人概念的基于博弈的仿真”，《未来交通前沿》，第3卷，2022年。[18] G. Zardini、N. Lanzetti、G. Belgioioso、C. Hartnik、S. Bolognani、F. D¨orfler与E. Frazzoli，“多模式出行系统中的战略互动：博弈论视角”，载于*IEEE*智能交通系统国际会议（*ITSC*），2023年，第5452–5459页。[19] D. Khulbe、R. Seshadri与M. Ben-Akiva，“评估纽约市拼车与拥堵收费影响的概率仿真框架”，《交通数据科学》，第5卷，第1期，第1–25页，2023年。[20] D. Gammelli、K. Yang、J. Harrison、F. Rodrigues、F. Pereira与M. Pavone，“面向可迁移自动驾驶按需出行的图元强化学习”，载于*ACM SIGKDD*知识发现与数据挖掘会议论文集，2022年，第2913–2923页。[21] B. Donovan与D. Work，“纽约市出租车行程数据（2010-2013年）”，2016年，伊利诺伊大学厄巴纳-香槟分校。[22] 美国人口普查局，“过去12个月的收入（表s1901），美国社区调查”，2013年，data.census.gov，访问日期：2026年1月13日。

附录

车辆再平衡模型

在每个时间步 t，模型都会确定运营商 o的再平衡流 $\{y^t_{i,j,o}\}$，在满足演员网络指定的期望车辆分布的同时，使总再平衡成本最小化。该优化问题的形式如下：

$$\min \quad \sum_{(i,j)\in\mathcal{E}} c^t_{i,j,o} y^t_{i,j,o} \tag{6}$$

$$\text{s.t.} \quad \sum_{j\neq i} (y^t_{j,i,o} - y^t_{i,j,o}) + m^t_{i,o} \geq \tilde{m}^t_{i,o}, \quad i \in \mathcal{V} \tag{7}$$

$$\sum_{j\neq i} y^t_{i,j,o} \leq m^t_{i,o}, \quad i \in \mathcal{V} \tag{8}$$

$$y^t_{i,j,o} \geq 0, \quad (i,j) \in \mathcal{E} \tag{9}$$

其中 $\tilde{m}^t_{i,\,o}$ 表示时间 t运营商 o 在区域i 的期望车辆数。目标函数 (6) 旨在最小化再平衡成本。约束条件 (7) 确保满足期望车辆数，同时考虑当前的闲置车辆$m^t_{i,\,o}$ 以及再平衡车辆的净流入。约束条件 (8) 根据各区域可用的闲置车辆数限制其再平衡流。期望车辆分布通过 $\tilde{m}^t_{i,\,o} = \lfloor w^t_{i,\,o} \sum_{i\in\mathcal{V}} m^t_{i,\,o} \rfloor$计算，其中 $w^t_{i,\,o}$ 是演员网络针对区域 i输出的再平衡权重。请注意，该网络流问题的约束矩阵是全酉模的，并且由于 $m^t_{i,\,o}$ 和 $\tilde{m}^t_{i,\,o}$ 均为整数，因此最优解必为整数。

选择模型敏感性

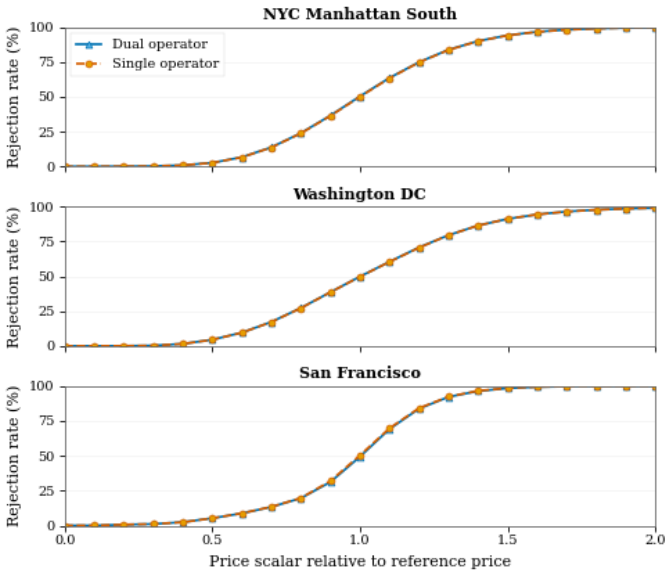


图 3：在所研究的数据集上，单运营商和双运营商设置中拒绝率随相对于历史参考价格的价格标量变化的情况；模型校准为在历史参考价格下拒绝率为 50%。

训练中使用的参数

类别	超参数	值
训练	行动者学习率（α_π）	2×10^{-4}
	评论家学习率（α_V）	4×10^{-4}
	折扣因子（γ）	0.97
	奖励缩放因子	4,000
	演员梯度裁剪	1,000
	评论家梯度裁剪	1,000
	评论家预热回合数	50
	训练回合数	150,000
网络	隐藏层大小	256
	前瞻时域（T）	6
	缩放因子	0.01
价格标量	观察起讫点价格	Yes
	起讫点价格标量	No

三步控制架构

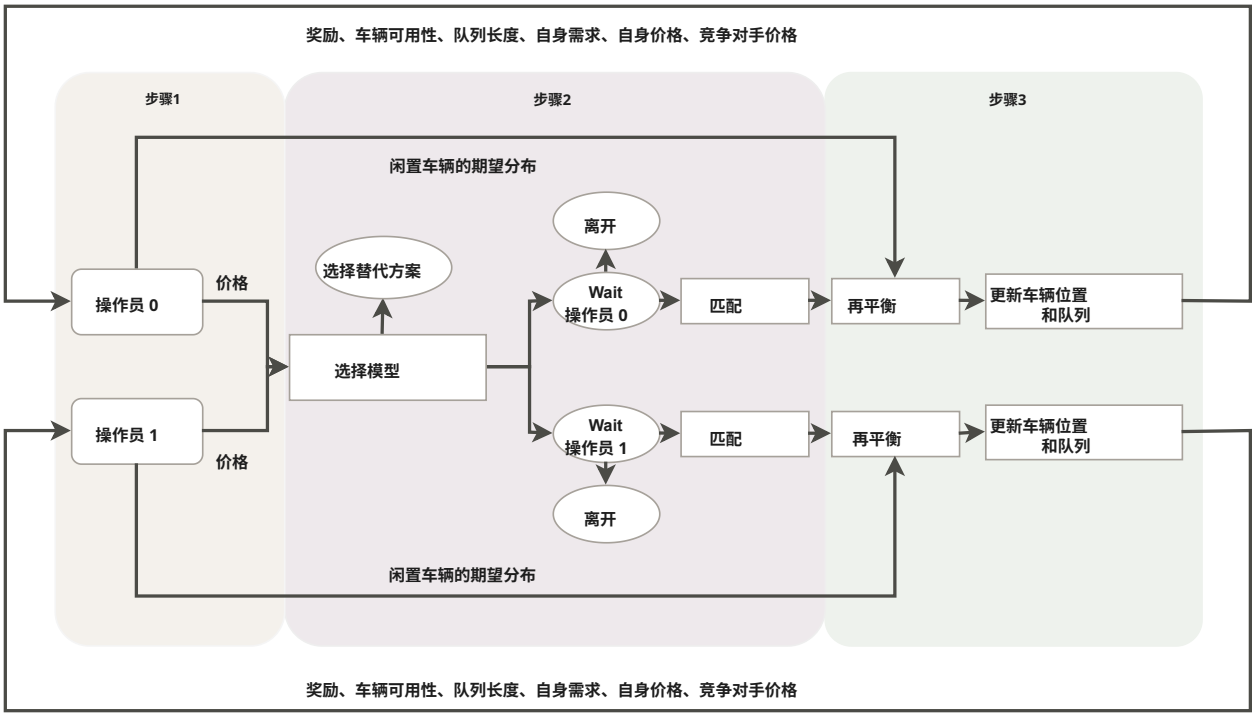


图4：用于双运营商自动驾驶按需出行控制的三步控制架构。步骤1：运营商制定定价和期望的空闲车辆分布策略。步骤2：通过选择模型、排队和匹配进行乘客分配。步骤3：进行空闲车辆再平衡，并更新车辆位置和队列。

运营商架构

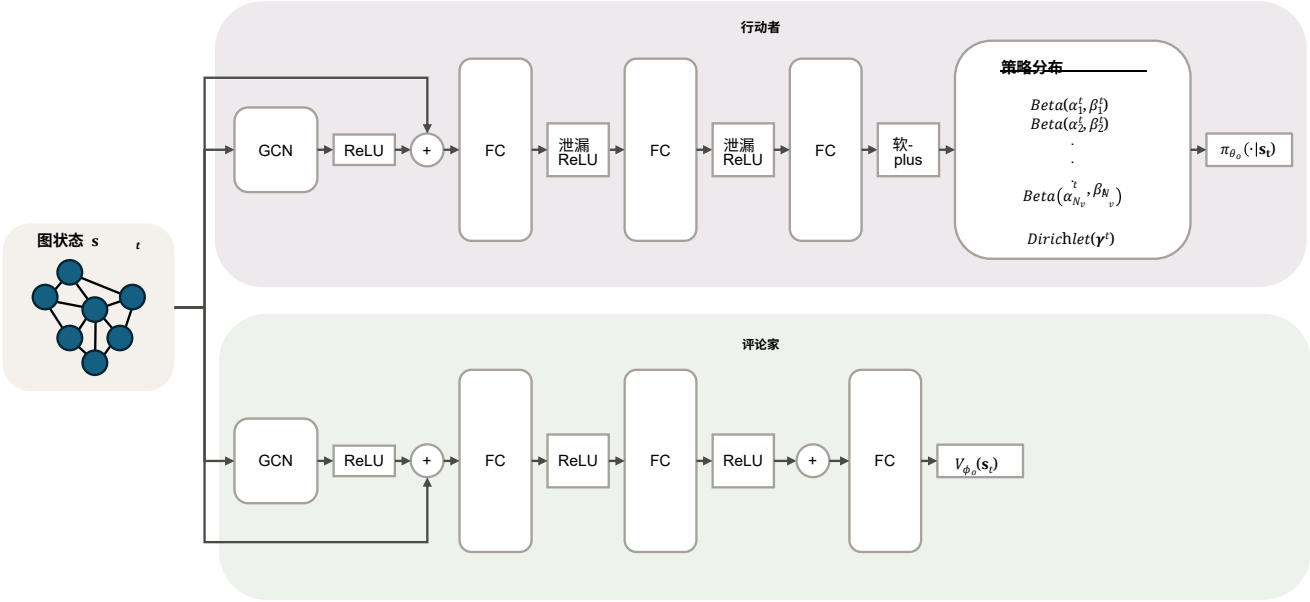
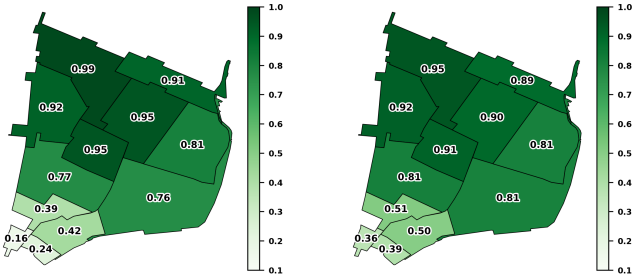


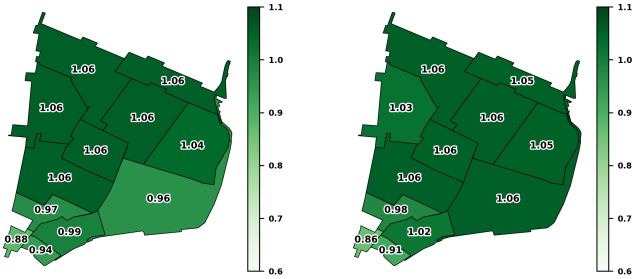
图5：运营商采用的演员-评论家架构。每个运营商都维护独立的行动者网络和评论家网络。

纽约市仅定价策略下的定价标量



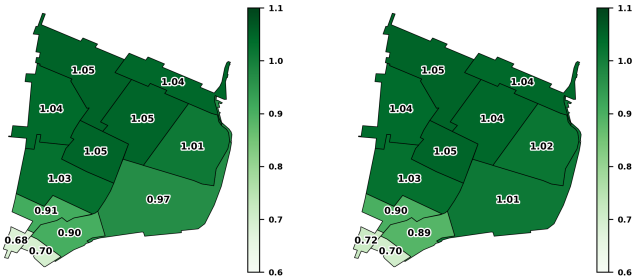
操作员 0　　操作员 1

图 6：仅定价策略在时间步 0 的初始定价策略。



操作员 0　　操作员 1

图 7：仅定价策略在时间步 19 的最终定价策略。



操作员 0　　操作员 1

图 8：仅定价策略在全部 20 个时间步上的时间平均定价标量。

车队规模敏感性

车队 Size	联合策略 奖励	联合策略 已服务	联合策略 再平衡成本	联合策略 再平衡行程	联合策略 价格 A0	联合策略 价格 A1	NC 奖励	NC 已服务	NC 价格 Both	UD 奖励	UD 已服务	UD 再平衡成本	UD 再平衡行程	UD 价格 Both
450	14170.5 (251.7)	2563.9 (28.4)	865.6 (66.9)	177.3 (16.0)	1.04 (0.00)	1.04 (0.00)	12270.7 (251.8)	2550.6 (15.9)	1.00 (0.00)	13217.4 (153.7)	2550.6 (15.9)	989.0 (70.0)	200.6 (16.6)	1.00 (0.00)
650	17481.1 (258.7)	3561.1 (29.0)	1379.0 (83.7)	284.7 (19.0)	0.96 (0.00)	0.96 (0.00)	16048.4 (416.6)	2894.2 (73.9)	1.00 (0.00)	17652.0 (277.5)	3496.5 (32.2)	1865.1 (98.3)	387.9 (19.6)	1.00 (0.00)
850	19513.1 (288.1)	4538.7 (38.9)	1871.0 (79.9)	386.3 (16.7)	0.88 (0.00)	0.88 (0.00)	18683.9 (461.8)	3367.4 (83.3)	1.00 (0.00)	19941.6 (334.8)	4171.5 (44.7)	3401.8 (128.6)	761.5 (27.2)	1.00 (0.00)
1050	20373.0 (198.7)	5512.9 (30.1)	2152.1 (77.1)	446.8 (19.7)	0.79 (0.00)	0.79 (0.00)	20667.8 (467.6)	3714.9 (83.3)	1.00 (0.00)	20450.0 (543.4)	4538.1 (72.0)	5016.9 (155.3)	1187.7 (41.2)	1.00 (0.00)
1250	20305.6 (235.2)	6440.1 (32.2)	2509.8 (117.9)	526.9 (29.9)	0.71 (0.00)	0.70 (0.00)	21955.0 (397.2)	3934.1 (71.2)	1.00 (0.00)	19798.1 (526.1)	4639.2 (77.2)	6259.8 (192.8)	1501.7 (46.1)	1.00 (0.00)

舰队拆分敏感性

舰队拆分 (O0:O1)	总计 奖励	O0 奖励	O1 奖励	总再平衡 行程	O0 再平衡 行程	O1 再平衡 行程	O0 价格	O1 价格	总计 已服务	O0 已服务	O1 已服务
5:5	17481.1 (258.7)	8717.0 (249.1)	8764.1 (223.5)	284.7 (19.0)	144.9 (17.8)	139.8 (14.3)	0.96 (0.00)	0.96 (0.00)	3561.1 (29.0)	1773.9 (26.1)	1787.2 (24.0)
4:6	17567.0 (231.0)	7295.2 (136.2)	10271.7 (243.7)	279.2 (17.8)	116.1 (12.6)	163.1 (14.6)	0.98 (0.00)	0.95 (0.00)	3577.2 (27.9)	1437.2 (14.7)	2140.0 (30.1)
3:7	17434.2 (241.7)	5772.6 (143.7)	11661.6 (259.2)	285.3 (19.8)	79.6 (11.5)	205.7 (18.9)	1.01 (0.00)	0.94 (0.00)	3562.5 (32.1)	1090.1 (17.4)	2472.4 (33.0)
2:8	16968.2 (263.3)	4140.0 (88.8)	12828.2 (244.2)	294.5 (21.2)	41.3 (4.9)	253.2 (19.8)	1.04 (0.00)	0.93 (0.00)	3530.7 (33.5)	742.3 (13.6)	2788.4 (31.9)
1:9	16655.4 (291.8)	2348.7 (77.9)	14306.7 (299.7)	311.6 (22.3)	14.1 (3.0)	297.5 (22.9)	1.10 (0.00)	0.93 (0.00)	3499.9 (35.8)	380.7 (8.7)	3119.2 (38.1)

定价信息的影响

	不共享信息			信息共享		
政策	Reb.	定价	联合	Reb.	定价	联合
总奖励	18096.6 (296.6)	18983.6 (267.4)	17689.4 (242.4)	18133.7 (330.8)	18631.6 (376.7)	17481.1 (258.7)
奖励算子 0	8981.1 (227.9)	9413.0 (111.9)	8951.6 (212.4)	9010.8 (242.6)	9351.6 (161.4)	8717.0 (249.1)
奖励算子 1	9115.5 (317.0)	9570.6 (192.9)	8737.8 (190.5)	9123.0 (341.4)	9280.1 (265.1)	8764.1 (223.5)
再平衡总成本	1544.4 (110.7)	—	1349.7 (76.5)	1600.2 (95.4)	—	1379.0 (83.7)
再平衡成本算子 0	806.0 (78.6)	—	662.5 (65.8)	813.5 (73.6)	—	699.6 (76.7)
操作员 1 再平衡成本	738.5 (103.5)	—	687.1 (59.2)	786.8 (107.0)	—	679.4 (63.1)
再平衡行程总数	327.8 (25.4)	—	276.5 (16.8)	338.4 (21.2)	—	284.7 (19.0)
操作员0的再平衡行程	172.0 (20.6)	—	134.7 (15.3)	173.5 (18.2)	—	144.9 (17.8)
运营商 1 再平衡行程	155.8 (23.6)	—	141.8 (12.4)	164.9 (24.2)	—	139.8 (14.3)
总服务需求	3521.9 (34.2)	3388.9 (38.1)	3579.5 (29.0)	3535.8 (39.4)	3337.6 (52.3)	3561.1 (29.0)
运营方0的服务需求量	1751.8 (26.1)	1678.1 (13.8)	1800.4 (25.4)	1758.1 (29.4)	1674.3 (19.1)	1773.9（26.1）
运营方1的服务需求量	1770.1（36.8）	1710.8（28.8）	1779.1（21.0）	1777.7（39.5）	1663.3 (38.7)	1787.2 (24.0)
价格算子 0	—	0.96 (0.00)	0.96 (0.00)	—	0.95 (0.00)	0.96 (0.00)
价格算子 1	—	0.95 (0.01)	0.96 (0.00)	—	0.95 (0.01)	0.96 (0.00)
等待时间（分钟）算子 0	1.82 (0.08)	1.34 (0.06)	1.97 (0.10)	1.80 (0.08)	1.51 (0.08)	2.03 (0.10)
等待时间（分钟）算子 1	1.81 (0.07)	1.37 (0.07)	2.06 (0.09)	1.78 (0.07)	1.45 (0.08)	2.06 (0.08)
队列算子 0	8.37 (1.09)	4.35 (0.70)	9.89 (1.12)	8.41 (1.12)	5.69 (0.73)	10.4 (1.3)
队列操作员 1	7.78 (1.21)	4.62 (0.80)	10.7 (1.4)	7.75 (1.05)	4.96 (0.67)	10.8 (1.2)
总需求	4683.0 (73.9)	4202.3 (55.2)	5185.7 (58.2)	4683.0 (73.9)	4311.7 (65.1)	5231.3 (61.6)
运营商 0 的需求	2336.2 (56.0)	2075.3 (34.3)	2552.0 (52.5)	2336.2 (56.0)	2145.4 (36.7)	2581.1 (59.5)
运营商 1 的需求	2346.8 (39.4)	2127.0 (31.8)	2633.7 (34.4)	2346.8 (39.4)	2166.3 (34.7)	2650.2 (42.1)

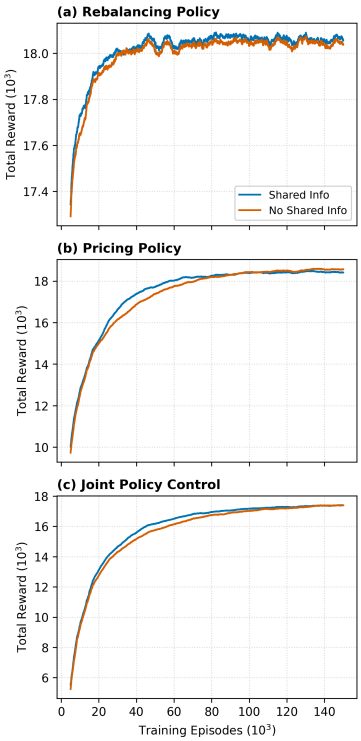


图9：对比有无竞争对手价格可见性场景的收敛动态：(a) 再平衡，(b) 定价，(c) 联合。曲线按 30 个回合进行平滑处理；为清晰起见，省略了前 5,000 个回合。请注意，这些奖励是基于从政策分布中采样得到的训练奖励，因此可能低于测试奖励。

区域工资异质性

指标	联合策略
总奖励	23564.6 (341.1)
奖励算子 0	11659.3 (206.4)
奖励算子 1	11905.3 (248.4)
再平衡总成本	2661.3 (87.2)
算子 0 的再平衡成本	1359.8 (52.2)
运营方1的再平衡成本	1301.5 (57.0)
再平衡行程总数	609.7 (21.0)
运营方0的再平衡行程数	315.0 (10.9)
运营方1的再平衡行程数	294.7 (15.0)
总服务需求	2990.9 (28.1)
运营方0的服务需求量	1480.6 (18.9)
运营方1的服务需求量	1510.3 (21.1)
价格算子 0	1.32 (0.00)
价格算子 1	1.31 (0.00)
等待时间（分钟）算子 0	1.42 (0.07)
等待时间（分钟）算子 1	1.46 (0.09)
队列算子 0	6.40 (1.01)
队列操作员 1	5.97 (1.34)
总需求	3567.0 (46.7)
运营商 0 的需求	1762.8 (24.5)
运营商 1 的需求	1804.2 (33.7)
平均工资	25.8 (0.1)

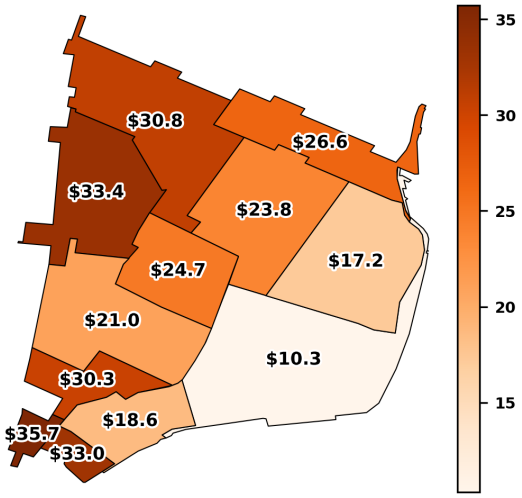


图 10：区域收入异质性下纽约市曼哈顿南部各区域乘客平均时薪分布。

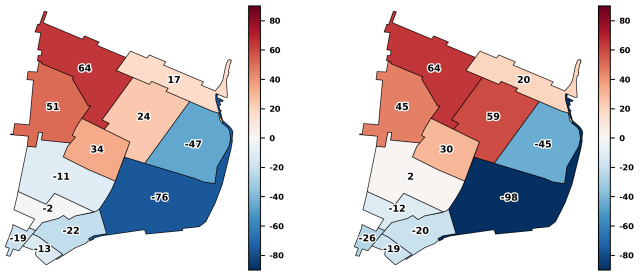


图 11：区域收入异质性下的净再平衡流，展示所有时间步的车辆累计移动情况。红色 =净接收方，蓝色 =净发送方。

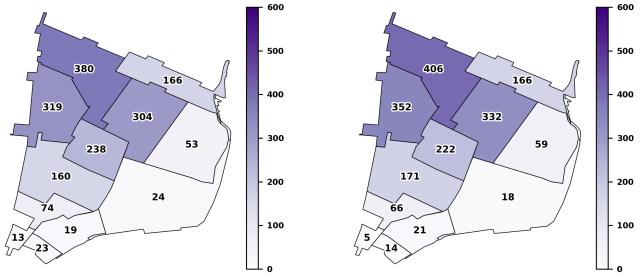


图 12：区域收入异质性下各区域产生的总需求。

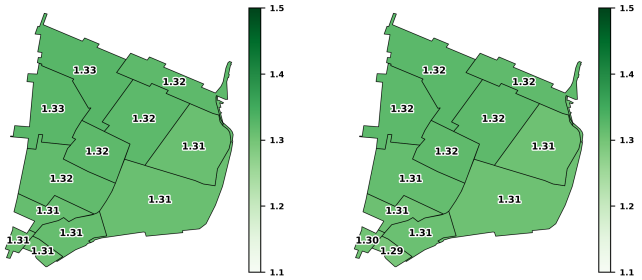


图 13：区域收入异质性条件下，各区域在所有时间步的平均定价标量。