

uber 与 lyft 共享出行科研前沿

2018.10.30 方建勇

1, 十年之内, 几乎所有大型汽车公司, 以及优步等车队运营商都宣布计划将自动驾驶汽车投入使用。与此同时, 除了提供减少碳足迹的好处之外, 电动汽车正迅速成为具有成本效益的下一代技术。集中管理的无人驾驶车辆车队以及电动车辆的运行特性相结合, 正在创造一种变革性的新技术, 可以在高服务水平下显着节省成本。这个问题涉及将车手分配给汽车的调度问题, 决定车队规模的计划问题, 以及决定每次旅行的价格的激增定价问题。在这项工作中, 我们建议使用近似动态规划来制定高质量的运营调度策略, 以确定哪辆车 (给定电池水平) 最适合特定行程 (考虑其长度和目的地), 何时应对车辆进行充电, 当它应该重新定位到提供更高密度旅行的不同区域时。然后, 我们使用自适应学习方法来讨论浪涌定价, 以确定每次旅行的价格。最后, 我们讨论了车队规模问题, 这取决于前两个问题。

2, 我们提出 Chorus, 一个具有新颖架构的系统, 用于为统计 SQL 查询提供差异隐私。我们的方法的关键是在执行之前将差异隐私机制嵌入查询中, 以便查询自动对其输出强制实施差异隐私。Chorus 与任何支持标准数学函数的 SQL 数据库兼容, 不需要用户修改数据库或查询, 同时支持许多差异隐私机制。据我们所知, 现有系统没有提供这些功能。我们使用四种通用差异隐私机制来演示我们的方法。在同

类型的第一次评估中，我们使用 Chorus 来评估真实世界查询和数据上的这四种机制。结果表明，我们的方法支持我们语料库中 93.9% 的统计查询，无需任何修改即可与生产 DBMS 集成，并可扩展到数亿条记录。Chorus 目前正在 Uber 部署其内部分析任务。Chorus 代表了公司 GDPR 合规工作的重要组成部分，可以提供差异化隐私和访问控制实施。在此职位上，Chorus 每天处理超过 10,000 个查询。

3, 了解城市规模的车辆移动性和出行模式对于解决从交通，污染到公共安全等许多问题至关重要。利用车辆移动性的时空分析，可以提出有希望的解决方案，以利用共享移动性和众包来缓解这些主要挑战。交通网络（例如 Uber, Lyft）的兴起仅仅是共享移动性的开始。在本文中，我们解决了旅行表示和匹配的问题。特别是，我们从空间和时间的角度研究了一个真实的旅行数据集（来自德国科隆）。对于依赖于时空现象的应用，期望比较轨迹。为此，我们提出了一种基于加权几何平均值（WGM）的新型组合时空相似性得分，并对其适用性和优势进行了实验。首先，我们使用得分来找到使用谱聚类在空间和/或时间上可分离的旅行群。然后将该分数用于 Catch-a-Ride (CaR) 和 CarPooling (CP) 场景的实时匹配。CaR 和 CP 分别实现了~40% 和~25% 的行驶距离减少，其代价是转移到下降位置（即驾驶员在接送时平均行驶距离<700m）-off for CP)。此外，在 CaR 情景中提供了与文献中可用指标的比较。我们发现 WGM 的主要优点包括支持时间或空间组件的灵活性，以及运行时复杂性的线性。最后，我们制

定了一个最佳的自由浮动汽车共享方案（例如，安排自动驾驶汽车或出租车系统），导致一小时内汽车平均服务 ≈ 3.88 次。

4，在基于流的编程中，数据源被抽象为可以通过回调函数操纵的值流。基于流的编程受到欢迎，因为它为处理交互式软件中的异步数据源提供了强大而富有表现力的范例。但是，高级流抽象也可能使开发人员难以推断其程序中的控制和数据流关系。当异步基于流的代码与线程受限的功能（例如限制对单个线程的 UI 访问的 UI 框架）交互时，这尤其具有影响力，因为流构造的线程行为通常是非直观的并且没有充分记录。在本文中，我们提出了一种基于类型的方法，可以静态地证明基于流的软件中 UI 访问的线程安全性。我们的主要观点是，流处理框架的流畅 API 使得能够通过类型细化来跟踪线程，从而可以自动推断运行代码的线程 - 一般来说是一个难题。我们将该系统实现为基于注释的 Java 类型检测程序，用于基于流行的 ReactiveX 框架构建的 Android 程序，并通过注释和分析 8 个开源应用程序来评估其功效，我们发现 33 个不安全的 UI 访问实例，同时仅产生注释负担每 186 个源代码行一个注释。我们还报告了我们将类型检测器应用于 Uber Technologies Inc. 代码库中两个更大的应用程序的经验，该应用程序目前在每次代码更改时运行，并阻止引入潜在线程错误的更改。

5，本文提出了一个通用的分析框架来模拟提供门到门服务的运输系统。该模型包括非共享出租车和需求响应运输（DRT）作为特殊情况。

在后一种情况下，我们既包括辅助乘客服务，如拨号乘坐（DAR），也包括目前由 Lyft 和 Uber 等众包出租车公司使用的共乘（共享出租车）形式。该框架产生了一些乐观的结果，因为除其他外，它是确定性的，并不跟踪跨太空的车辆。然而，由于其简单性，该框架产生了许多感兴趣的情况的近似闭合形式公式。

6, 近年来，城市环境中的强大导航已经获得了相当大的学术和商业利益。这主要是由于谷歌和优步等大型商业组织进入自主导航市场。大多数研究都避开了基于全球导航卫星系统（GNSS）的导航。对利用 GNSS 数据的厌恶是由于城市环境中数据的性质下降（例如，多径，卫星可见性差）。城市环境中 GNSS 数据的退化使得传统（GNSS）定位方法（例如，扩展卡尔曼滤波器，粒子滤波器）表现不佳。然而，主要应用于基于同时定位和映射（SLAM）的机器人应用的基于鲁棒图理论的传感器融合方法的最新进展也可以应用于 GNSS 数据处理。本文将利用一种称为因子图的方法，结合几种稳健的优化技术来评估它们对强大的 GNSS 数据处理的适用性。这项研究的目标是双重的。首先，对于 GNSS 应用，我们将在图论理论估计框架内通过实验评估鲁棒优化技术的有效性。其次，通过发布用于本研究的软件和数据集，我们将为佐治亚理工学院平滑和映射（GTSAM）库引入一个新的开源前端，以便整合 GNSS 伪距观测。

7, 人的生命很重要。允许在我们的道路上行驶的自动驾驶车辆的决

定非常重要。这一直是政策制定者，技术专家和公共安全机构之间争论的热门话题。最近优步公司（Uber Inc.）的自驾车撞车导致行人死亡，这加剧了自动驾驶汽车技术尚未准备好在公共道路上部署的论点。在这项工作中，我们分析了优步车祸并揭示了一个问题：“Uber 车祸是否可以避免？”。我们将最先进的计算机视觉模型应用于这种高度实用的场景。更一般地，我们的实验结果是对各种图像增强和物体识别技术的评估，以使用优步碰撞作为案例研究在低照明条件下实现行人安全。

8, 我们提出了一种模型，用于优化从配送中心到最终接收者的 n 个包裹的最后一英里交付，使用结合使用乘坐共享平台（例如 Uber 或 Lyft）和传统内部货车交付的策略系统。主要目标是计算为 n 个包中的每个包提供给私人驱动程序的奖励，使得递送所有包的总预期成本最小化。我们的技术方法基于离散顺序包装问题的制定，其中在区间 $[0, T]$ 期间随机时间从仓库中拾取捆包。我们的理论结果包括精确和渐近（如 $n \rightarrow \infty$ ）表达式，用于预期数量的包将被时间 T 拾取，并且与经典的 Renyi 的停车/打包问题密切相关。

9, 各种汽车和移动公司，例如福特，优步和 Waymo，目前正在公共道路上测试他们的预制自动驾驶车辆（AV）车队。然而，由于安全关键案例的罕见性以及实际上无限数量的可能交通场景，这些道路测试工作被认为是乏味，昂贵且具有风险的。在本研究中，我们提出加

速部署框架，以安全有效地估算公共街道上的 AV 表现。我们通过适当地解决逐步准确性改进和自适应地选择部署 AV 的有意义和安全的环境表明，所提出的框架产生高度准确的估计，更快的评估时间，更重要的是，降低部署风险。我们的研究结果为目前激烈而积极的讨论如何在公共道路上正确测试 AV 性能提供了答案，从而为 AV 技术实现安全，高效和统计可靠的测试框架

10, 我们探讨了 Lyft 和 Uber 等乘坐共享服务的动态供需问题, 这些服务需求随时间和地理位置而波动。我们寻求最大化社会福利, 这取决于出租车和乘客位置, 乘客对服务的估值以及出租车和乘客之间的距离。我们唯一的控制手段是设定激增价格, 然后出租车和乘客根据这些价格最大化其公用事业。我们研究了两个相关的模型: 类似于 Wardrop 模型的连续乘客 - 出租车设置, 以及离散的乘客 - 出租车设置。在连续的环境中, 每个位置都被一组无限小的战略出租车和一组无限小的非战略性乘客所占据。在离散设置中, 每个位置都由一组战略代理人, 出租车和乘客占用, 乘客具有不同的服务价值。我们将连续模型扩展为与时间相关的设置并研究相应的在线环境。如果存在使得出租车大概移动的最小成本流量 (a) 每个出租车都遵循最佳响应, (b) v 值高于 v 的激增价格 r_v 的所有战略乘客, 则喘振价格处于乘客 - 出租车均衡中。服务和 (c) 没有服务价值低于 r_v 的战略乘客 (非战略无限小乘客总是服务)。本文计算了激增价格, 使得由此产生的乘客 - 出租车均衡最大化社会福利, 并且这种激增价格的计算

是在聚合时间内。此外，它是乘客揭示其真实价值的主导策略。我们力求在网络环境中实现社会福利最大化，并为此目的提供紧密的竞争比率。我们的在线算法利用随时间和地理位置计算的激增价格，引发连续的乘客 - 出租车平衡。

11, 随着具有 GPS 功能的移动设备的广泛使用, 可以从各种来源获得前所未有的轨迹数据, 例如 Bikely, GPS-wayPoints 和 Uber。创新运输服务的兴起和最近自动驾驶汽车的突破将导致轨迹数据和相关应用的持续增长。在新兴平台中支持这些服务将需要在轨迹数据库中进行更有效的查询处理。在本文中, 我们提出了两个新的轨迹数据库覆盖查询: (i) k 最大化对轨迹的反向距离搜索 ($kMaxRRST$); (ii) 轨迹上的最大 k 覆盖范围搜索 ($MaxkCovRST$)。我们提出了一种新颖的索引结构, 即弹道四叉树 (TQ-tree), 它利用四叉树将轨迹分层组织成不同的四叉树节点, 然后应用 z 排序, 通过每个节点内的空间局部进一步组织轨迹。这种结构在修剪轨迹搜索空间方面非常有效, 这是独立的兴趣。通过利用 TQ 树数据结构, 我们开发了一种分而治之的方法来计算轨迹“服务价值”, 并采用最佳策略来探索轨迹, 使用适当的服务价值上限来有效地处理 $kMaxRRST$ 查询。此外, 为了解决 $MaxkCovRST$, 这是一个非子模块的 NP 难问题, 我们提出了一个贪婪的近似, 它也利用了 TQ 树。我们通过对几个真实数据集的广泛实验研究来评估我们的算法, 并证明我们的基于 TQ 树的算法优于普通基线两到三个数量级。

12, 随着许多在线和离线服务, 个人 (出租车司机) 的顺序匹配 (供应和需求) 系统 (超级, Lyft, Grab 为出租车; ubereats, deliveryoo 等用于食品; amazon prime, lazada 等用于杂货) 的出现, 送货员, 送货车司机等) 通过在“正确”的时间在“正确”的地方赚取更多。我们专注于学习技术, 以便在其他“学习”个体面前为个人提供指导 (在正确的时间适当的位置)。个体之间的相互作用是匿名的, 即交互的结果 (竞争需求) 与代理的身份无关, 因此我们将这些称为匿名 MARL 设置。现有的相关性研究是使用强化学习 (RL) 或多智能体强化学习 (MARL) 进行独立学习。聚合系统中的个体数量非常大, 个人有自己的私利 (最大化收入)。因此, 传统的 MARL 方法要么不具有可扩展性, 要么共同目标或行动协调的假设不可行。在本文中, 我们专注于通过利用匿名性来提高独立强化学习者的表现, 特别是流行的深度 Q 网络 (DQN) 和 Advantage Actor Critic (A2C) 方法。具体而言, 我们使用代理密度分布的熵来控制由其他代理引入的非平稳性。我们展示了个人和所有代理人我们的学习者在聚合系统和现实出租车数据集的通用实验设置上的收入的显著改善

13, 在过去十年中, 我们称之为数字媒体工作场所的增长。数字媒介工作场所是利益相关者之间的互动主要由专有的, 算法管理的数字平台管理的工作场所。平台利益相关者之间的关系取代是这些工作场所的一个关键特征, 也是每个利益相关者彼此之间合同责任减少的一个

促成因素。在本文中，我们讨论了这种结构和缺乏问责制作为数字媒介工作场所 Uber 共享应用程序 Uber 实现偏差的根源或至少是推动者的一些方式。

14, 交叉路口是危险的地方。威胁来自行人，自行车和车辆之间的相互作用，没有车道标记的更复杂的车辆轨迹，阻止了解谁拥有通行权的阶段，隐形车辆接近，车辆障碍物和非法行动。 Vision Zero 计划中规定的“道路饮食”和道路重新设计并未完全解决这些挑战。它们的车载传感器以及对传感器读数的不懈关注也不会被自动驾驶汽车完全克服。事故也可能发生，因为司机，骑自行车者和行人没有他们需要的信息来避免错误的决定。在这些情况下，可以通过智能交叉点计算和传达丢失的信息。该信息给出当前的全信号相位，相位将改变的时间的估计，驾驶员或自动车辆的盲点的占用以及红灯违反者的检测。本文开发了智能交叉口的设计，其动机是在亚利桑那州坦佩的一个交叉路口，在自动优步沃尔沃和手动本田 CRV 之间分析事故。智能交叉点用作使用 I2V 通信的“受保护的交叉点”。

15, 城市化率的不断提高给我们社区已经受到限制的交通网络带来了压力。 Uber 和 Lyft 等乘坐共享平台正变得越来越普遍，特别是在城市环境中。虽然这些服务由于其按需性质而被认为比乘坐公共交通更方便，但报告显示它们并不一定能减少主要城市的拥堵。关键问题之一是通常移动性决策支持系统关注个体效用并且仅在拥塞出现后作

出反应。在本文中，我们提出了社会考虑的多模态路由算法，这些算法是主动的，并通过预测考虑乘客对移动服务整体效能的共同影响。我们调整了 MATSim 模拟器框架以结合所提出的算法，对田纳西州纳什维尔的案例研究进行了模拟分析，评估了我们的路由模型对不同渗透水平和采用社会考虑路线的交通拥堵的影响。我们的结果表明，即使在低渗透率（社会比率）下，我们也能够实现系统级性能的提升。

16, 出租车需求预测是智能城市智能交通系统的重要组成部分。准确的预测模型可以帮助城市预先分配资源以满足旅行需求，并减少街道上的空车，这会浪费能源并加剧交通拥堵。随着优步和滴滴出行（中国）等出租车服务的日益普及，我们能够不断收集大规模的出租车需求数据。如何利用这些大数据来改进需求预测是一个有趣且关键的现实问题。传统的需求预测方法主要依赖于时间序列预测技术，这些技术无法模拟复杂的非线性空间和时间关系。深度学习的最新进展通过学习复杂特征和大规模数据的相关性，在传统上具有挑战性的任务

（如图像分类）方面表现出色。这一突破激发了研究人员探索交通预测问题的深度学习技术。然而，现有的业务量预测方法仅考虑了空间关系（例如，使用 CNN）或时间关系（例如，使用 LSTM）。我们提出了一个深度多视图时空网络（DMVST-Net）框架来模拟空间和时间关系。具体而言，我们提出的模型包括三个视图：时间视图（通过 LSTM 建模未来需求值与近时间点之间的相关性），空间视图（通过本地 CNN 建模局部空间相关性）和语义视图（共享相似时间的区域

之间的建模相关性) 模式)。对大型真实出租车需求数据的实验证明了我们的方法相对于最先进的方法的有效性。

17, 共乘经济正在经历快速增长和创新。Uber 和 Lyft 等公司继续以相当大的速度增长, 同时提供他们的平台作为共享服务的组织媒介, 增加消费者效用以及雇用数千名兼职职位。然而, 在共乘服务的建模方面仍然存在许多挑战, 其中许多目前尚未得到广泛的考虑。在本文中, 开发了一个基于代理的模型来模拟华盛顿特区大都市区的共乘服务。该模型用于检查通用共乘服务的乘客(客户)和驾驶员(服务提供商)获得的效用水平。提供了个人代理 Metro-Washington Area Ridesharing Model (IAMWARM) 的描述, 以及典型模拟运行的描述。我们在两种情景和两种空间运动行为下调查了 24 小时驾驶员的财务收益。我们描述了两种空间行为是随机运动和 Voronoi 运动。两种运动行为都在固定运行条件情景和可变运行条件情景下进行了测试。我们发现 Voronoi 运动增加了驾驶员获得的效用, 但是这种系统属性的出现仅在变化的情景条件下才可行。这一结果提供了两个重要的见解: 第一个是乘客取件之前的驾驶员运动决定会影响服务和驾驶员的经济收益, 从而影响乘客成功获得的比率。第二, 这种现象仅在实施乘客和驾驶员到达率变化的实验条件下才是明显的。

18, 训练现代深度学习模型需要大量计算, 通常由 GPU 提供。将计算从一个 GPU 缩放到多个可以实现更快的培训和研究进度, 但需要

两个复杂性。首先，培训库必须支持 GPU 间通信。根据所采用的特定方法，这种通信可能需要从可忽略不计到显着的开销。其次，用户必须修改他或她的训练代码以利用 GPU 间通信。根据培训库的 API，所需的修改可能是重要的或最小的。在 TensorFlow 库下启用多 GPU 培训的现有方法需要不可忽略的通信开销，并要求用户大量修改他们的模型构建代码，导致许多研究人员避免整个混乱并坚持使用较慢的单 GPU 培训。在本文中，我们介绍了 Horovod，这是一个改进了缩放障碍的开源库：它通过减少环来实现高效的 GPU 间通信，并且只需对用户代码进行几行修改，从而在 TensorFlow 中实现更快，更简单的分布式培训。Horovod 可通过 <https://github.com/uber/horovod> 上的 Apache 2.0 许可证获得

19, 电子招聘服务的需求正在快速增长，特别是在大城市。优步是美国和纽约市第一家也是最受欢迎的电子商务公司。2014 年和 2015 年纽约市黄色出租车和优步的需求比较表明，对优步的需求有所增加。然而，这种需求可能不是在空间上或时间上均匀分布的。使用时空时间序列模型可以帮助我们更好地理解对电子外包服务的需求并更准确地预测它。本文分析了一个时间模型（向量自回归（VAR））和两个时空模型（时空自回归（STAR）；最小绝对收缩和选择算子应用于 STAR（LASSO-STAR））和不同的预测性能场景（基于时间和空间滞后的数量），并适用于高峰时段和非高峰时段。结果表明需要考虑出租车需求的时空模型。

20, 我们提出了 Al / Al₂O₃ 界面的粘合剂和结构特性的计算研究, 作为金属 - 绝缘体 - 金属 (MIM) 隧道器件的构建块, 其中电子传输通过隧道机制通过夹层绝缘屏障完成。本文的主要目的是在原子尺度上理解几何细节在隧道屏障剖面形成中的作用。为了提供可靠的结果, 我们仔细评估了用于检查 Al / Al₂O₃ 界面的传统方法的准确性。这些是最广泛使用的交换相关函数, LDA, PBE 和 PW91, 用于预测平衡界面距离和粘附能量的通用结合能关系 (UBER), 以及作为结稳定性度量的理想分离工作。最后, 我们对每个连接处的原子和面间弛豫进行了详细分析。我们的结果表明, Al 膜表面的结构不规则性对降低隧道势垒高度有显著贡献, 而 Al 膜中的晶面间弛豫远离直接界面对隧道性质没有显著影响。另一方面, 在隧道势垒的成形中可能涉及多达 5-7 层的 Al₂O₃。这些层的晶面间弛豫取决于界面的几何形状, 并且可导致相对于块状氧化物中的相应厚度的净收缩 13%。这是一个重要的数量, 因为隧道概率在屏障宽度上呈指数关系。

21, 近年来, 亚马逊市场, AirBnB, Uber / Lyft 和 Upwork 等许多成功的电子商务市场平台崛起, 其中一个中央平台调解买卖双方之间的经济交易。在这些平台的推动下, 我们制定了一套设施位置问题, 我们称之为双面设施位置。在我们的模型中, 代理到达基础度量空间中的节点, 其中任何买方和卖方之间的度量距离捕获相应匹配的质量。平台在节点上发布价格和工资, 并打开一组设施以将代理路由到。假

设任何设施的代理都匹配。该平台通过在节点与其路由到的设施之间施加距离约束来确保高匹配质量。它通过确保设施的流量至少是预先指定的下限来确保高服务可用性。受这些限制，平台的目标是在预算平衡较弱的情况下最大化社会盈余（或贸易收益），即利润是非负的。我们提出了一个针对这个问题的近似算法，它对任何常数 $\epsilon > 0$ 产生过剩的 $(1 + \epsilon)$ 近似，同时通过常数因子放宽匹配质量（即任何匹配的最大距离）。我们使用 LP 舍入框架，可以轻松扩展到其他目标，例如最大化交易量或利润。我们通过考虑动态市场设置来证明我们的模型，其中代理根据随机过程到达并且具有匹配的有限耐心（或期限）。我们执行排队分析以显示对于将代理路由到设施并与之匹配的策略，确保代理的放弃概率降低以确保到达每个设施的足够流量。

22, 我们是否可以从大型用户生成的数据集中有效地提取有用信息，同时保护用户的隐私和/或确保表示的公平性。我们将此问题转换为删除稳健的子模块最大化的实例，其中部分数据可能由于隐私问题或公平性标准而被删除。我们提出了第一种具有内存效率的集中式，流式传输和分布式方法，它们具有针对任意数量的对抗性删除的恒定因子近似保证。我们广泛地评估了我们的算法在现实世界应用程序中的先进技术性能，包括 (i) 具有位置隐私限制的 Uber-picking 位置；(ii) 特征选择，公平性限制收入预测和犯罪率预测；(iii) 对人口普查数据的删除摘要具有鲁棒性，包括 2,458,285 个特征向量。

23, 许多“共享经济”平台, 如优步和 Airbnb, 已越来越受欢迎, 为消费者提供更多选择, 供应商有机会获利。然而, 他们也带来了有关监管, 税收义务和对城市环境影响的新问题, 并引起了各种利益集团的激烈辩论。关于这些问题的实证研究是有限的, 部分原因是相关数据不可用。在这里, 我们的目标是了解共享经济的服务提供商, 调查谁加入和谁受益, 使用美国的 Airbnb 市场作为案例研究。我们将 188,000 多家拥有的超过 211,000 个爱彼迎房源与人口统计, 社会经济地位 (SES), 住房和旅游特征联系起来。我们表明, 无论参与形式或年份如何, 收入和教育始终是加入 Airbnb 的两个最有影响力的因素。家庭收入中位数较低的地区, 或拥有学士学位和更高学历的居民比例较高的地区往往拥有更多的寄宿家庭。然而, 在考虑上市表现时, 按新收到的评论数量衡量, 我们发现收入对整个房屋上市有积极影响; 位于家庭收入中位数较高的地区的房源往往会有更多的新评论。我们的研究结果从经验上证明, SES 弱势地区的劣势和 SES 优势领域的优势可能存在于共享经济中。

24, 时间序列预测的可靠不确定性估计在许多领域都至关重要, 包括物理学, 生物学和制造业。在优步, 概率时间序列预测用于强大预测特殊事件期间的出行次数, 驾驶员激励分配以及跨数百万度量的实时异常检测。经典时间序列模型通常与用于不确定性估计的概率公式一起使用。但是, 这些模型难以调整, 扩展和添加外生变量。受近期长期短期内存网络复苏的推动, 我们提出了一种新颖的端到端贝叶斯深

度模型，该模型提供时间序列预测和不确定性估计。我们提供了所提出的完整行程数据解决方案的详细实验，并成功应用于优步的大规模时间序列异常检测。

25, 当前的旅行需求模型无法预测旅行行为的长期趋势，因为它们不需要建立新的交通方式（Uber, Lyft 等）的会员资格和市场份额的机制。我们建议整合离散选择和技术采用模型以解决上述问题。为此，我们建立了离散混合模型的制定，特别是潜在类选择模型（LCCM），它们与网络效应模型集成在一起。网络效应模型量化了新技术的空间/网络效应对采用效用的影响。我们采用了一种确认方法，根据技术扩散文献的结果来估算我们的动态 LCCM，这些文献侧重于定义两种不同类型的采用者：创新者/早期采用者和模仿者。LCCM 允许各种细分市场采用的异质性，即创新者/早期采用者，模仿者和非采用者。我们利用来自美国主要城市的单向汽车共享系统的显示偏好（RP）时间序列数据来估计模型参数。这些数据需要在推出后的 2.5 年内为汽车共享服务提供一整套会员注册。与技术扩散文献一致，我们的模型确定了三个潜在类别，其采用的效用具有明确定义的一组重要且行为一致的偏好。技术采用模型预测某个人在某个时间段采用该服务的概率，并通过社会影响，网络效应，社会人口统计和服务水平属性来解释。最后，对模型进行了校准，然后用于预测汽车共享系统的采用情况，以用于潜在的投资策略方案

26, Uber 和 Didi 等租赁平台在全球越来越受欢迎。然而, 利用共享经济的未经授权的采购活动可能会严重损害这一新兴产业的健康发展。作为规范按需乘坐服务和消除黑市的第一步, 我们设计了一种基于轨迹从汽车池中检测出租车的方法。由于法律问题, 一些城市可能无法公开获得许可的外包汽车痕迹并且可能完全缺失, 我们转而将公共交通开放数据 (即出租车和公共汽车) 的知识转移到普通车辆的检测中。我们提出了一个两阶段转移学习框架。在第 1 阶段, 我们将出租车和公交车数据作为输入, 使用出租车/公共汽车和其他车辆共享的轨迹特征来学习随机森林 (RF) 分类器。然后, 我们使用 RF 标记所有候选汽车。在第 2 阶段, 利用前一阶段高可信标签的子集作为输入, 我们进一步学习用于 Ridesourcing 检测的卷积神经网络 (CNN) 分类器, 并通过 co 迭代地改进 RF 和 CNN 以及特征集。- 训练过程。最后, 我们使用由此产生的 RF 和 CNN 集合来识别候选池中的 ridesourcing 汽车。真实汽车, 出租车和公共汽车轨迹的实验表明, 我们的转移学习框架, 不需要预先标记的租赁数据集, 可以达到与监督学习方法类似的准确性。

27, 在本文中, 我们研究了著名的 k-server 问题的随机变体。在 k-server 问题中, 我们需要最小化服务于度量中的 t 个请求的在线序列的 k 个服务器的总移动。在随机设置中, 我们预先给予独立分布 $\langle P_1, P_2, \dots, P_t \rangle$, 并且在每个时刻步骤 i 从 P_i 提取请求。在这种设置中设计最优在线算法是 NP 难的, 因此我们的工作重点是设计近似最优的在线算

法。我们首先展示了某类非自适应在线算法的结构特征。我们证明，在一般指标中，最好的这种算法的成本不会低于最优在线算法的三倍。接下来，我们提出一个整数程序，为任意度量找到该类的最优算法。最后，通过舍入该程序的线性松弛的解，我们提出了随机 k -服务器问题的在线算法，其中线和度量中的近似因子为 3，而一般度量的度量中的 $O(\log n) \cdot \tilde{n}$ 。此外，我们定义优步问题，其中每个需求由两个端点组成，即源和目标。我们证明了给定 k -服务器问题的近似算法，我们可以得到 Uber 问题的 $(a + 2) -$ 近似算法。事实上，需求通常与我们研究随机优步问题的时间高度相关。此外，我们将结果扩展到相关设置，其中请求到达某一点的概率不仅取决于时间步长，还取决于先前到达的请求。

28, 最近，平台生态系统作为一个关键的商业概念受到关注。平台用户通过彼此提供和/或要求内容来实现平台生态系统的可持续增长：例如，Facebook, YouTube 或 Twitter。由于平台所有者使用和销售其业务数据，因此平台生态系统中用户数据的重要性和价值更加突出。对于不同利益相关者之间的数据滥用或滥用，隐私问题和收益分享，人们越来越担心。传统的数据治理侧重于通用目标和管理企业数据的通用方法。它需要对平台生态系统的复杂情况和关系提供有限的支持，其中多个参与方贡献，使用数据和分享利润。本文通过文献综述确定平台生态系统的数据治理因素。该研究随后调查了四个平台生态系统的数据治理实践状态：Facebook, YouTube, EBay 和优步。最后，将

工业和学术界的 19 个治理模型与我们为平台生态系统确定的数据治理因素进行比较，以揭示差距和局限性。

29, 电网是人类有史以来建造的最大, 最复杂的基础设施之一。现代文明依赖于工业生产, 人类流动和舒适生活。然而, 诸如 60 赫兹变压器等许多关键技术是在 20 世纪初开发出来的, 并且从那时起变化很小。¹ 通过输电 - 配电网从发电到客户的传统单向电力也在名义上发生了变化, 但它不再满足 21 世纪市场能源客户的需求。一方面, 1.28 亿美国住宅用户每月支付 15 亿美元的水电费, 但他们没有选择能源供应商的选择。在许多传统产业被数字互联网技术 (亚马逊, Ebay, 优步, Airbnb) 转变的世界中, 传统的电能市场明显滞后。需要迈向真正的数字网格。这样的数字网格需要物理层 (能量和功率) 与数字和网络信息的紧密集成, 以允许类似于电子商务世界的开放和实时市场。推动这种激进转型的另一个主要因素是能源资源所有权和负荷流量的快速变化模式。在分布式太阳能, 能源存储, 电动汽车, 现场发电和微电网的成本下降的推动下, 分布式能源 (DER) 的高渗透率从控制的角度将挑战大大转移到电网边缘。设想的数字网格必须促进开放竞争和开放式创新, 以加速采用新的 DER 技术, 同时满足电网稳定性, 数据爆炸和网络安全方面的挑战。

30, GPS 使移动设备能够不断提供改善日常生活的新机会。例如, Uber 或公共交通管理局创建的应用程序中收集的数据可用于规划运输路

线, 估算容量, 并主动识别低覆盖区域。在本文中, 我们研究了一种新的查询 - 反向 k 最近邻搜索 (RkNNT), 可用于路径规划和容量估计。给定一组现有路线 DR , 一组乘客过渡点 DT 和查询路线 Q , RkNNT 查询返回将 Q 作为其最近 k 个行驶路线之一的所有过渡。为了解决这个问题, 我们首先开发一个索引来处理动态轨迹更新, 以便最新的转换数据可用于回答 RkNNT 查询。然后, 我们引入了一个过滤器细化框架, 用于使用建议的索引处理 RkNNT 查询。接下来, 我们将展示如何使用 RkNNT 来解决最佳路线规划问题 MaxRkNNT (MinRkNNT), 即搜索从起始位置到最终位置的最佳路线, 该路线可以吸引最大 (或最小) 乘客数量。预定义的行程距离阈值。真实数据集上的实验证明了我们方法的效率和可扩展性。据我们所知, 这是研究 RkNNT 路线规划问题的第一项工作。

31, 随着现代交通系统变得越来越复杂, 需要允许旅行者在城市中有有效导航的移动应用程序。在出租车运输中, 最近 Uber 的扩散引入了新的规范, 包括灵活的定价方案, 其中旅程成本可以根据乘客需求和驾驶员供应而快速变化。为了在最适合的旅程中做出明智的选择, 旅行者需要实时获取有关提供商定价的知识。为此, 我们开发了 OpenStreetcab 移动应用程序, 提供出租车运输方面的建议, 比较供应商价格。我们描述了它在伦敦和纽约两个城市的开发和部署, 并分析了数千个用户旅程查询, 以比较优步与当地主要出租车供应商的价格模式。我们观察到两个城市的出租车运输市场存在巨大的异质性。这

促使我们在地面上进行价格验证和测量实验，比较伦敦的 Uber 和 Black Cabs。实验结果揭示了有趣的见解：他们不仅确认了专业司机用户收到的价格和服务质量的反馈，而且还揭示了出租车提供商之间价格和旅程时间之间的权衡。特别是在旅行时间方面，我们展示了与依赖现代导航系统的驾驶员相比，在大多数情况下，经验丰富的出租车司机能够更快地驾驶到目的地。我们提供的证据表明，在城市密度高的城市中心，这种优势变得更加强大。

32，出租车服务和产品交付服务对我们现代社会起着重要作用。由于共享经济的出现，Uber，Didi，Lyft 和 Google 的 Waze Rider 等乘坐共享服务正变得越来越普遍，并成为我们日常生活中不可或缺的一部分。然而，这些服务的效率受到供需之间的次优和不平衡匹配的严重限制。我们需要一个通用的框架和相应的有效算法来解决有效匹配问题，从而优化这些市场的性能。现有的出租车和送货服务研究仅适用于单边市场的情况。相比之下，这项工作研究了市场经济中的出租车和交付服务（缩写为“出租车和交付市场”）的高度概括模型，可广泛用于双边市场。此外，我们为不同的应用提供有效的在线和离线算法。我们在实际设置下通过理论分析和跟踪驱动模拟来验证我们的算法。

33，这项工作提出了一个新的框架，通过基于社交媒体数据的自动内容分析的开放式创新方法开发运输中的新产品和服务。该框架能够从在线社交网络（OSN）中提取用户评论，通过信息提取和情感分析技

术处理和分析文本，以获得有关市场上产品接收的相关信息。使用 Uber 移动应用程序开发了一个用例，Uber 是当今世界上发展最快的技术公司之一。我们测量了一个有争议的，高度分散的事件如何影响关于优步的推文数量和用户的看法。虽然优步的形象没有变化，但是观察到该公司的推文数量大幅增加，这意味着其产品的自由和重要的传播。

34，出租车是纽约市公共交通系统的一个重要方面。在曼哈顿无处不在的黄色出租车与城市的地铁系统一样具有标志性，近年来，该市引入了绿色出租车，以改善中央商务区和机场以外地区的出租车服务。每天约有 500,000 次出租车旅行，载有约 80 万名乘客，不包括 Uber，Lyft 或 Carmel 等其他制服公司。自 2008 年以来，黄色出租车已经能够使用信用卡处理票价，而信用卡在票价总额中的份额越来越大。然而，使用信用卡支付出租车费用在各街区之间差异很大，并且出租车费用的现金支付与无银行账户或欠银行账户的人口之间存在很强的相关性。这些问题引起政策制定者的关注，因为该市约有 10% 的家庭没有银行账户，而在一些社区，无银行账户的家庭比例超过 50%。在本文中，我们使用多个数据集来探索邻域的出租车费用，并研究如何使用出租车服务与使用传统银行服务相关联。对于乘客支付现金的倾向存在明显的空间维度，我们发现移民身份和“无银行账户”都是出租车现金交易的强有力预测因素。这些结果对于租用车辆行业的当地法规有影响，特别是在需要信用卡的服务快速增长的背景下。如果没有

某种类型的现金支付选项，出租车服务将隔离某些社区。至少，现有和新的过境服务提供商必须考虑获取主流金融产品，作为其股权分析的一部分。

35, 本文考虑了一个闭路排队网络模型的共乘系统，如滴滴出行，Lyft 和优步。我们专注于空车路线，这是一种机制，通过该机制我们控制网络中的车流以优化系统范围的实用功能，例如，乘客到达时空车的可用性。我们将排队网络的流程级和稳态收敛建立到大型市场体系中的流体限制，其中对汽车的乘坐和供应的需求倾向于无限，并且使用该限制来研究基于流体的优化问题。我们证明了基于流体的优化获得的最优网络效用是有限汽车系统中用于任何静态和动态路由策略的效用的上限，在该策略下闭合排队网络具有静态分布。在基于流体的最优路由策略下渐近地实现该上限。滴滴出行发布的真实数据的仿真结果证明了与其他各种政策相比，使用基于流体的最优路由策略的好处

36, 使用片上网络（NoC）接近理想的线路延迟是许多核心系统（尤其是数百个核心）的重要实际问题。虽然其他研究人员专注于优化大型网格，绕过或推测路由器流水线，或创建更复杂的对数拓扑，但本文提出了一种平衡组合，可以简化队列缓冲。使用执行驱动仿真对 PARSEC 和 SPLASH 的 9 个基准进行初步分析表明，当将每个网状端口单个核心连接到至少 50% 时，利用率从 2% 上升到 2%，因为集中

器和路由器队列的延迟时间比使用高出约 6 倍。仅 20 个周期的理想延迟。也就是说，16 端口网格就足够了，因为排队是系统性能的不常见情况。通过这种方式，网格跳数被限制为 3，因为负载通过扩展的浓度变得均匀，并且使用用于网状路由器的传统四级流水线以及较小的对数边缘来逼近理想等待时间。一个逼真的 Uber 也很详细，具有与使用优化路由器管道的 64 端口网格相同的性能，将基线提高了 12 %。正在进行的工作开发了通过调整缓存块的位置来更好地平衡负载的技术，并将 Uber 与无缓冲路由进行比较。

37, 平面图上的动态最短路径问题要求我们预处理平面图 G ，这样我们可以支持 G 中边的插入和删除以及任意两个节点 u, v 之间的距离查询，受制于图形保持平面的约束。一直。在过去的几十年里，这个问题在理论界和实验界都得到了广泛的研究，并且每天都被谷歌，微软和优步等公司解决了数百万次。最著名的算法基于 Fakcharoenphol 和 Rao [FOCS'01] 的开创性论文的想法，在 $O(\sqrt{n})$ 时间内执行查询和更新。Abraham 等人的 $(1+\epsilon)$ -近似算法。[STOC'12] 在 $O(\sqrt{n})$ 时间内执行更新和查询。具有 $O(\text{polylog}(n))$ 运行时的算法将是一项重大突破。然而，这样的运行时间仅对于模型中的 $(1+\epsilon)$ -近似是已知的，其中由于 Abraham 等人仅允许限制的权重更新。[SODA'16]，或者更容易解决连接等问题。在本文中，我们遵循最近和非常活跃的工作，基于流行的猜想显示多项式时间问题的下界，获得平面图中自然问题的第一个这样的结果。由于已知减少的高度非平

面性质和“平面化小配件”的不可能性，这些结果以前是遥不可及的。我们引入了一个新的框架，其灵感来自文献中关于距离标记方案和参数化复杂性的技术。使用我们的框架，我们表明，除非传统的 APSP 问题可以解决，否则平面图中的动态最短路径或最大权重二分匹配的算法都不能支持在分摊的 $O(n^{12-\epsilon})$ 时间内的更新和查询， $\epsilon > 0$ 。在真正的亚临时时间， [...]

38，共享经济的增长是由共享平台的出现推动的，例如 Uber 和 Lyft，它们希望与想要租用它们的客户分享他们的资源。这种平台的设计是经济学和工程学的复杂混合，如何“优化”设计这样的平台仍然是一个悬而未决的问题。在本文中，我们将重点放在共享平台的价格和补贴设计上。我们的结果提供了对收入最大化价格和社会福利最大化价格之间权衡的见解。具体而言，我们引入了一种新的共享平台模型，并描述了该模型中利润和社会福利最大化的价格。此外，我们将效率损失限制在利润最大化的价格下，表明在实际环境中利润与效率之间存在强烈的一致性。我们的结果强调，由于供应短缺，平台的收入可能在实践中受到限制；因此，平台有强烈的动机鼓励通过补贴分享。我们提供了这种补贴何时有价值的分析特征，并说明如何优化所提供补贴的规模。最后，我们使用来自中国最大的共乘平台滴滴出行的数据，验证了我们分析的见解。

39，今天的云服务越来越多地使用其他运行服务的功能构建。在本文

中，我们质疑传统的服务质量（QoS）指标和执行技术是否足够，因为它们是以生产者为中心的。我们认为，与银行系统和许多共享经济应用（例如 Uber 和 Airbnb）中的客户评级系统类似，应引入消费质量（QoC）以捕获有关服务消费者的不同指标。我们展示了消费者和提供商如何将 QoS 和 QoC 的组合（称为 QoX）用于改善其基础架构的安全性和管理。此外，我们还演示了如何在其他消费者和提供商之间共享信息，从而提高 QoX 的价值。为了解决共享信息（即 sybil 攻击和错误信息）的主要挑战，我们描述了如何利用云提供商作为担保机构来确保信息的完整性。我们提出了在云环境中对适当的抽象和接口进行原型设计的初步结果，重点关注设计对服务提供商和消费者的影响。

40，优步最近在城市出租车运输方面引入了新的做法。旅程价格几乎可以实时动态变化，并且在城市中从一个地区到另一个地区的地理位置也会发生变化，这种策略称为激增定价。在本文中，我们探索新一代开放数据集的力量，以了解公共交通领域出现的新破坏技术的影响。我们的主要目标是为城市上班族提供更透明的经济环境，我们提供优步与纽约黄色出租车公司之间的直接价格比较。我们发现优步尽管标准定价较低，但实际上平均收费较高，特别是在短期内，但频繁发生，出租车旅行。基于这种洞察力，我们开发了一个智能手机应用程序 OpenStreetCab，为移动用户提供个性化的咨询，出租车提供商的旅程便宜。该应用程序推出近五个月后，在一个城市吸引了超过三千名用

户。他们的旅程查询为类似技术可能为城市上班族带来的潜在节省提供了额外的见解，其中一个亮点是，如果他们选择最便宜的出租车服务提供商，纽约的用户平均每辆出租车节省 6 美元。我们进行了大量实验，以显示优步的激增定价是更高旅程价格的驱动因素，从而为我们的应用用户带来更高的潜在节省。最后，由于观察到优步的激增定价更频繁地出现在直觉上，我们制定了一个预测任务，其目标是预测地理区域的激增趋势。使用外部到优步数据集，我们展示了如何估算一个区域内的客户需求，以及通过扩展激增定价，高精度。

41，优步作为全球替代出租车运营商的崛起最近吸引了很多人的兴趣。除了讨论新现象的媒体头条，例如关于如何扰乱传统的交通运输业，决策者，经济学家，公民和科学家进行了一场围绕将新一代共享经济服务融入城市生态系统的手段进行的讨论。在这项工作中，我们的目标是通过利用描述纽约市黄色出租车移动性的公开纵向数据集，为讨论提供新的视角。除了移动之外，该数据还包含出租车客户为每次旅行支付的票价信息。因此，我们有机会在世界上最大的大都市中心之一提供标志性的黄色出租车与其现代竞争对手优步之间的首次比较。我们确定了优步 X，优步版本的优步出租车服务，往往比同一旅程中的黄色出租车更贵。我们还展示了优步的经济模型如何有效地利用人类运动中众所周知的模式。最后，我们通过提出一个新的移动应用程序来进一步分析我们的分析，该应用程序比较城市的出租车价格，以方便旅行者的出租车选择，希望最终能够降低通勤成本。我们的研究

提供了一个案例,说明公共大型数据集如何通过提供缺乏最新法规的经济部门透明度的机会来改善消费者的城市服务。