山 东 科 技 大 学

本科毕业设计（论文）外文译文

题 目 共享单车用户行为视角下的上海市骑行轨迹与通 勤模式分析

学院（系、部） 测绘与空间信息学院

专业班级 地理信息科学20级2班

姓 名 潘 志 清

学 号 202001020717

指 导 教 师 牟 乃 夏

填表时间： 2024 年 4 月 16 日

**填写说明**

1. 外文译文需添加外文原文内容和外文译文内容。
2. 字体使用中文宋体小4号，英文Times New Roman 小4号，间距段前段后均为0行，行距固定值20磅，首行缩进2字符，并用A4纸打印，于左侧装订成册。

**摘要**

自行车作为一种环保的交通方式已经被世界各地的许多政府所推广。因此，建设有效的自行车道已经成为推广骑行生活方式的政府的重要任务，因为规划良好的自行车道可以减少交通拥堵，降低骑行者和机动车驾驶员的安全风险。不幸的是，现有的自行车道规划轨迹挖掘方法并未考虑到关键的现实政府约束：1）预算限制，2）施工便利性，和3）自行车道利用率。

在本文中，我们提出了一种基于大规模真实世界自行车轨迹数据开发自行车道建设计划的数据驱动方法。我们强制执行这些约束条件来制定我们的问题，并引入了一个灵活的目标函数来调节用户数量的覆盖范围和其轨迹长度之间的收益。我们证明了该问题的NP-难度，并提出了基于贪婪算法的启发式方法来解决它。最后，我们在Microsoft Azure上部署了我们的系统，提供了广泛的实验和案例研究，以展示我们方法的有效性。

**引言**

自行车作为日常通勤常用的城市交通方式，已经被世界各地的多个政府所推广，原因有几点：1）对用户来说，它是一种价格实惠且环保的交通方式；2）它减少了道路交通拥堵；3）它是一种健康的生活方式。因此，建设有效的自行车道（如图1a所示）已经成为政府推广骑行生活方式的重要任务。规划良好并实施得当的自行车道不仅使骑行更加便利，还减少了骑行者和机动车驾驶员的安全风险。

传统的城市自行车道规划方法主要依赖于经验和调查。随着GPS嵌入式设备的普及，基于数据驱动的自行车道规划方法也逐渐出现，例如，现有的一些研究仅仅聚焦于总结自行车轨迹数据的共性，而忽略了政府面临的现实约束和需求：

预算限制。在道路段上实现自行车道存在成本，其中可能包括：1）为创建自行车道而需要的空间；和2）建造自行车道栏杆和绘制标志的费用（如图1(a)所示）。不幸的是，政府经常预算有限。

施工便利性。为了实施自行车道，需要向施工区派遣施工队，所需的队伍数量也是一种硬性约束。为了便于管理，政府希望避免将团队分散到偏远地区的施工区域（例如，图1(b)中红色线条突出显示了具有最多自行车轨迹的前100个段），而是更倾向于将它们聚集在一起，即在道路网络中作为有限数量的连通组件。

自行车道利用率。从政府的角度来看，建造自行车道的目标是增加更多骑行者的可用性并覆盖更多可能的路线，作为公共服务。

为了纳入这些现实约束，本文提出了一种基于从摩拜1用户收集的大量轨迹的数据驱动方法，用于规划自行车道。摩拜是一种完全无桩的共享单车系统，目前在中国的许多大城市部署。它是世界上最大的自行车运营商，并最近使上海成为了世界上最大的共享单车城市。与传统的基于站点的共享单车系统相比，摩拜用户生成的轨迹在解决自行车道规划问题时具有两个独特的优势：

真实的出行需求。与许多现有的基于站点的共享单车系统不同，这些系统要求用户从指定的站点取车和还车，摩拜提供了一个更灵活的系统，用户可以在任意位置取车和还车（如图2(a)所示）。因此，摩拜用户的轨迹反映了实际的城市出行需求。

丰富的出行信息。摩拜的锁系统上嵌入了一个3G通信组件和一个GPS模块（如图2(b)所示），使用户可以通过手机找到自行车。它还记录了用户所经过的确切路线（如图2(c)所示），而传统的基于站点的共享单车系统只能提供用户的登记/退车信息。

在本文中，我们设计、实现并部署了一个基于Microsoft Azure的数据驱动自行车道规划系统，该系统不仅利用了数千名摩拜用户生成的大量自行车轨迹，还满足了政府提出的约束和目标。所提出的系统包含两个主要组件：1）预处理，用于预处理来自摩拜用户的轨迹，并将其映射到道路网络上；2）自行车道规划，该组件接受用户输入（即政府的要求）并提供自行车道建议。主要贡献总结如下：

我们通过考虑各种施工约束来形式化自行车道规划问题，并提出了一个灵活的调节参数，以表征覆盖用户数量和连续覆盖自行车行程长度之间的设计权衡。该问题证明是NP-难的。

我们提出了一种贪婪的网络扩展算法，为基于数据驱动的自行车道规划问题提供了可扩展且近似的解决方案。为了获得更好的效果，我们还提出了两种不同的算法初始化方法，分别适用于低预算和高预算情况。

我们对来自上海市的一个月摩拜轨迹数据（即2016年9月1日至2016年9月30日）进行了广泛评估。我们还进行了广泛的数据分析，并发现了许多有用的见解。此外，我们进行了现场案例研究，以评估我们自行车道建议的有效性。

我们部署了一个包含真实数据集的在线系统，并在Microsoft Azure上提供了该系统。最后，我们收集了政府官员的反馈，我们的系统收到了非常积极的评价。

论文的其余部分组织如下：第2节描述了问题定义和系统概述。第3节介绍了预处理模块。第4节介绍了自行车道规划模块。实验和案例研究在第5节给出。第6节介绍了系统部署细节和专家评价。相关工作在第7节总结。第8节总结了论文。

**相关工作**

在本节中，我们总结了三个主要领域的相关工作：1）基于数据驱动的城市规划，2）轨迹数据挖掘，和3）传统的自行车道规划方法。

基于数据驱动的城市规划。随着用户、车辆和公共交通系统大量移动数据的可用性，城市计算技术在许多城市规划任务中变得越来越受欢迎，因为大量的移动数据反映了现实世界中的真实出行需求。例如，[39]挖掘出租车轨迹中的模式，以建议道路建设和公共交通项目。[35]根据交通模式和POI分布推断出城市中不同的功能区。[5, 17]根据多个移动数据集在城市中识别潜在的交通模式和异常情况。在本文中，我们专注于提供一种基于数据驱动的方法，以寻找更有效和经济的自行车道规划方式。

轨迹数据挖掘。自行车道规划问题与轨迹数据挖掘相关。许多系统已经提出，基于大量轨迹数据发现经常使用的路线，例如，有些项目聚类/总结道路网络上的轨迹，这有助于城市规划者了解热门路线并改进公共交通系统。最接近的自行车轨迹挖掘项目是[10, 11, 19]，它们专注于总结轨迹的共性，并找出自行车道的K-Primary走廊。然而，所有这些工作都不能直接用于自行车道规划，因为它们未考虑到现实中的预算和连接约束。

传统的自行车道规划。城市中的传统自行车道规划主要在交通领域进行研究，并且严重依赖经验经验，例如，为了评估建设自行车道的必要性，[29, 32]基于公共调查和地理统计数据提供了一些高层建议，如道路网络和POI分布。一些尝试[18]基于调查数据系统地发现实际自行车路线选择的因素。最近，还有一些基于基于站点的共享单车系统的交通预测和路线建议的工作，例如，[23, 25]。

**结论**

在本文中，我们提出了一种基于摩拜（一种主要的无桩共享单车系统）在上海市收集的真实自行车轨迹的数据驱动方法来规划自行车道。我们的系统可以以更加现实的方式解决自行车道规划问题，考虑到城市规划者的约束和要求：1）预算限制，2）施工便利性，和3）自行车道利用率。我们还提出了一个灵活的有益评分函数，用于调整覆盖用户数量和覆盖行程长度之间的偏好。我们证明了所制定的问题是NP-难的，因此我们提出了一种贪婪的网络扩展算法，并提供了两种不同的初始化方法：基于top-k和空间聚类。

我们对大规模的摩拜数据进行了大量实验，并展示了我们提出的自行车道规划框架的有效性。我们观察到了有趣的权衡现象，即基于top-k（或空间聚类）的初始化方法在低（或高）建设预算下表现良好。我们还进行了基于我们路径推荐结果的现场案例研究，并提出了许多重要见解，以提高特定地区的骑行便利性。一个演示系统已经在Microsoft Azure上部署供公众使用，来自上海市徐汇区政府官员的专家反馈确认了我们系统的有效性和可用性。

最后，在未来的工作中，我们计划利用Microsoft Azure中的并行计算框架来提高系统响应时间，以更高效地处理更大的轨迹数据集。此外，我们希望进一步探索交互式规划过程，以融入更多人类智能。

**ABSTRACT**

Cycling as a green transportation mode has been promoted by many governments all over the world. As a result, constructing effective bike lanes has become a crucial task for governments promoting the cycling life style, as well-planned bike paths can reduce traffic congestion and decrease safety risks for both cyclists and motor vehicle drivers. Unfortunately, existing trajectory mining approaches for bike lane planning do not consider key realistic government constraints: 1) budget limitations, 2) construction convenience, and 3) bike lane utilization.

In this paper, we propose a data-driven approach to develop bike lane construction plans based on large-scale real world bike trajectory data. We enforce these constraints to formulate our problem and introduce a flexible objective function to tune the benefit between coverage of the number of users and the length of their trajectories. We prove the NP-hardness of the problem and propose greedy-based heuristics to address it. Finally, we deploy our system on Microsoft Azure, providing extensive experiments and case studies to demonstrate the effectiveness of our approach.

**INTRODUCTION**

Cycling as a commonly used urban transit mode for daily commute has been promoted by multiple governments all over the world [1, 40] for several reasons: 1) it is an affordable and environment-friendly transportation mode for users; 2) it reduces road traffic congestion; and 3) it is a healthy lifestyle [31]. As a result, building effective bike lanes, demonstrated in Figure 1a, becomes a vital task for governments to promote the cycling lifestyle. Well planned & implemented bike lanes not only make cycling easier, but also reduce the safety risks for both cyclists and drivers of motor vehicles [30].

Traditional approaches to planning bike lanes in a city rely mainly on empirical experience and surveys [12, 18, 32]. With widespread availability of GPS embedded devices, more data-driven approaches on planning bike lanes have emerged, e.g., [10, 11, 19]. However, existing works [10, 11, 19] merely focus on summarizing commonalities of bike trajectory data while ignoring the realistic constraints and requirements faced by the government:

Budget Limitations. There are costs to realizing a bike lane on a road segment, which may include: 1) the space for creating bike lanes; and 2) the price of building bike lane railing, and painting signs (demonstrated in Figure 1(a)). Unfortunately, governments often have limited budgets.

Construction Convenience. To implement the bike lanes, construction teams need to be dispatched to construction zones, with the number of teams required also being a hard constraint. For the sake of ease of management, the government would like to avoid spreading teams out to construction zones in far reaching locations (e.g., red lines in Figure 1(b) highlights the top-100 segments with the most bike trajectories), and prefer to have them clustered, i.e., as a limited number of connected components in the road network.

Bike Lane Utilization. As a public service, from the government’s point of view, the objective of building bike lanes is to increase the usability for more bikers and cover more possible routes.

To incorporate these real world constraints, in this paper, we propose a data-driven approach for planning the bike lanes based on the massive number of trajectories collected from Mobike 1 users. Mobike is a fully station-less bike-sharing system currently deployed in many large cities in China. It is the world’s largest bike operator, and recently made Shanghai the world’s largest bikeshare city. Compared to the traditional station-based bike sharing system, trajectories generated by Mobike users have two distinctive advantages in tackling the bike lane planning problem:

Realistic Travel Demands. Unlike many existing station-based bike sharing systems, which require the users to pick up and drop off bikes from designated stations, Mobike offers a more flexible system, where the users can pick up and drop off their bikes at arbitrary locations (Figure 2(a)). As a result, the trajectories of Mobike users reject actual urban travel demands.

Rich Travel Information. A 3G communication component and a GPS module are embedded on the lock system in Mobike (demonstrated in Figure 2(b)), which enables the users to find bikes with their phones. It also keeps the track of the exact route traversed by the users (Figure 2(c)), while the traditional station-based bike sharing system can only provide the check-in/out information.

In this paper, we design, implement and deploy a data-driven bike lane planning system on Microsoft Azure, which not only leverages the massive bike trajectories generated by thousands of Mobike users, but also fulfills the constraints and objectives requested by the government. The proposed system contains two main components: 1) Pre-Processing, which pre-processes the trajectories from the Mobike user and maps them on the road network; and 2) Bike Lane Planning, which takes the user’s input (i.e., requirements from the government) and provides bike lane suggestions. The main contributions are summarized as follows:

We formulate the bike lane planning problem by considering various construction constraints, and propose a flexible tuning parameter to characterize the design trade-off between the number of covered users and the length of the continuously covered bike trips. The problem proves to be NP-hard.

We propose a greedy network expansion algorithm, which provides a scalable and approximate solution to the data-driven bike lane planning problem. To achieve a better effectiveness, we also propose two different approaches to initialize the algorithm, which work well for low and high budget scenarios, respectively.

We evaluate the proposed algorithms extensively over one month Mobike trajectory data (i.e., from 9/1/2016 - 9/30/2016) from the City of Shanghai. We also provide an extensive data analysis and discover many useful insights. Moreover, on-field case studies are conducted to evaluate the effectiveness of our bike lane recommendations.

An online system with the real dataset is deployed and available on Microsoft Azure [2]. Finally, we collect the feedback from the government officials, from which our system received very positive reviews.

The rest of the paper is organized as follows: Section 2 describes the problem definition and the system overview. Section 3 presents the pre-processing module. Bike lane planning module is presented in Section 4. Experiments and case studies are given in Section 5. Section 6 presents the system deployment details and the expert reviews. Related works are summarized in Section 7. Section 8 concludes the paper.

**RELATED WORK**

In this section, we summarize the related works in three main areas: 1) data-driven urban planning, 2) trajectory data mining, and 3) traditional bike lane planning methods.

Data-Driven Urban Planning. With the availability of massive amounts mobility data from users, vehicles and public transportation systems, urban computing techniques have become more and more popular in many urban planning tasks, as the massive mobility data rejects real travel demands in the physical world [38]. For example, [39] mines patterns in taxi trajectories to suggest road constructions and public transportation projects. [35] infers different function zones in a city based on traffic patterns and POI distribution. [5, 17] identify potential traffic patterns and anomalies in the city based on multiple mobility datasets. In this paper, we focus on providing a data-driven approach to find a more effective and economic way for bike lane planning.

Trajectory Data Mining. The bike lane planning problem is related to the trajectory data mining [6, 16, 21, 22, 24, 26, 27, 34]. Many systems have been proposed to discover frequently used routes based on massive trajectory data, e.g., [6, 15, 16, 21, 26, 27]. There are also some projects on clustering/summarizing trajectories on the road network [13, 20], which help urban planners to know the popular routes and improve public transportation system. The closest projects on bike trajectory mining are [10, 11, 19], which focus on summarizing the trajectory commonality and find out the K-Primary Corridors for bike lanes. However, all of these works can not be directly used for bike lane planning, as they fail to consider the realistic budget and connectivity constraints.

Traditional Bike Lane Planning. Traditional bike lane planning in a city is mainly studied in the transportation domain, and relies heavily on the empirical experience, e.g., [8, 12]. To evaluate the necessity of building bike lanes, [29, 32] provide some high level suggestions based on public surveys and the geographical statistics, such as the road network and POI distributions. There have been some attempts [18] to systematically discover factors for actual bike route choices based on survey data. Recently, there have also been some works on traffic predication and route suggestion based on the station-based bike-sharing systems, e.g., [23, 25].

**CONCLUSION**

In this paper, we propose a data driven approach to plan bike lanes based on the real bike trajectories collected from Mobike (a major station-less bike sharing system) in the City of Shanghai. Our system can address the bike lanes planning problem in a more realistic way, considering the constraints and requirements from urban planners’ perspective: 1) budget limitations, 2) construction convenience, and 3) bike lane utilization. We also propose a flexible beneficial score function to adjust preferences between the number of covered users and the length of covered trips. The formulated problem is proven to be NP-hard, thus we propose a greedy network expansion algorithm with two different initialization methods: top-k based and spatial clustering.

We perform extensive experiments on a large scale Mobike data and demonstrate the effectiveness of our proposed bike lane planning framework, where interesting trade-off phenomena are observed namely the top-k based (resp. spatial clustering based) initialization approach works well with low (resp. high) construction budgets. We also conduct an on-field case study based on our path recommendation results, and present many important insights to improve cycling convenience in a given area. A demonstration system is deployed on Microsoft Azure for public use, and the expert feedback from the government officials from Xuhui District, Shanghai, confirms the effectiveness and usability of our system.

Finally, in future work, we plan to use the parallel computing framework in Microsoft Azure to improve system response time to work more efficiently with larger trajectory datasets. Also, we would like to further explore the interactive planning process to incorporate more human intelligence.