

中国科学技术大学

学士学位论文



数据驱动的电动车充电站 布置方法

姓 名:	姜 昊 茗
院 系:	少年班学院
学 号:	PB13000422
导 师:	刘恒昌 副研究员
完成时间:	二〇一七年二月

University of Science and Technology of China
A dissertation for bachelor's degree



Data Driven Approach for Deploying Charging Station for Electric Vehicles

Author :	<u>Haoming Jiang</u>
Department :	<u>The School of Gifted Young</u>
Student ID :	<u>PB13000422</u>
Supervisor :	<u>Henchang Liu</u>
Finished Time :	<u>February, 2017</u>

目 录

目 录	I
摘 要	III
ABSTRACT	V
第一章 引言	1
第二章 文献综述	3
2.1 电动车充电站的布置	3
2.1.1 传统做法	3
2.1.2 多因素综合决策	3
2.1.3 大数据驱动的办法	4
2.2 轨迹数据挖掘	4
2.2.1 数据预处理	5
2.2.2 数据管理	5
2.2.3 数据查询	6
2.2.4 挖掘任务	7
2.2.5 隐私保护	8
2.2.6 轨迹数据挖掘的应用	8
第三章 数据驱动的充电桩布置办法	11
3.1 问题定义	11
3.2 数据描述	12
3.3 解决方案	13
3.3.1 地图分块	13
3.3.2 子轨迹抽取	14
3.3.3 需求核密度估计	14
3.3.4 最短距离期望解	15
3.4 实验结果	16
3.4.1 可视化分析	16
3.4.2 期望最短寻找距离	17
第四章 总结	19
4.1 总结	19
4.2 未来工作	19

参考文献.....	21
致 谢.....	27

摘 要

在工业化的现代，面对传统化石燃料带来的诸多问题，人们对清洁能源的需求与日俱增。电动车诸如特斯拉，比亚迪也越来越被大众使用。因此相应的配套设施的要求将成为城市规划重要的一环。本文主要研究了如何在城市里设计充电站的排布方案。本文详细的叙述了国内外相关领域现在已有的相关研究工作，为将来的研究工作提供了清晰的参考。主要以充电站布置和轨迹数据挖掘方面的内容为主。

当前是数据的时代，本文也探讨了如何利用大数据进行建模。有别于传统的纯粹优化的方案，数据驱动的途径提供了解决这个问题的新思路。特别的利用了深圳出租车的历史行驶数据和深圳市的现有地图数据 (包括道路，已有充电站位置)，估计了充电的需求。根据充电需求，建立了核密度估计模型和 Top-k 模型，寻找最优的充电站位置配备方案。

关键词： 电动车，充电站，优化，排布，大数据

ABSTRACT

Our society is highly industrialized. When facing the problems of fossil fuel, people now focusing on the clean energy more than ever. Electrical Vehicles such as Tesla and BYD, are now process an increasing portion of vehicle market. As a result related infrastructures are now is very important to civil engineering. This article study how to make optimal charging stations deployment for a certain city. This article exhaustively investigated related research work. It provide a clear references for future research work. Most contents are related to Charging Station Deployment and Trajectory Data Mining.

Since this is the era of big data, this article also discussed how to use the big data to establish the model. It is different to the traditional optimization way. Data driven approach provide a novel thinking to this problem. Particularly, it utilized the Shenzhen taxi historical trajectory data and Shenzhen map data (including roads and existing charging stations) to estimate the charging requirement and the optimal traveling strategy in the city network. Corresponding to the charging demand, by using Top-k model and kernel density estimation model the optimal deployment of charging station can be found.

Keywords: Electrical Vehicles, Charging Station, Optimization, Deployment, Big Data

第一章 引言

随着科技的发展，人们对生活质量的要求不再仅仅是效率。健康环境问题成为了大家目前最关心的一个问题。清洁能源也慢慢被大家所重视。汽车在作为人类最重要的出行交通工具之一的同时，也是一个重要的污染源。化石燃料燃烧后所带来的排放问题直接或间接地导致了雾霾的产生。因此利用电力这样的清洁能源汽车被大众和国家所关注。相关企业诸如比亚迪，特斯拉都一直在进行相关的研发。他们的产品也逐步逐步被大众所接纳。比如在深圳市就引进了大量的电动出租车和清洁能源公交车。与之伴随而来的一个问题便是，不断增长的电动汽车带来不断增长的电力需求。电动车充电桩的建设将成为市政建设的一个重要的组成部分。

现如今是一个大数据的时代，数据的获取和处理的代价都被大大降低。数据驱动的方法和应用如雨后春笋般不断地涌现出来。现如今 GPS 设备是非常常见的，通过收集车辆上装备的 GPS 设备产生的数据，我们可以建立许多的相关应用。既然现如今有为电动车建设充电桩的需求，大数据必然能够从中发挥它的重要作用。如何处理车辆的行驶轨迹数据，建立模型解决得到最优的充电桩布置方案就成为了一个问题。

本文首先对已有的相关文献做了详尽的调研。其中主要包括两个方面：**1.充电桩排布方案 2.轨迹数据挖掘**。充电桩的排布方案，国内外已经有了非常多的探讨。有基于优化，模拟的办法。而且这个问题类似于加油站的建立排布问题，是属于传统的优化工业工程领域的重点研究问题之一。过去的研究对现在新出现的充电站排布问题有很好的参考价值。轨迹数据挖掘方面，由于轨迹数据具有数据量大，难以处理等特点，相关的研究也非常丰富。从最基本的预处理，储存，基本查询，构建上层应用都有丰富研究值得参考。详尽的文献综述不仅是对自己阅历的一个丰富，也为将来的研究提供了参考。

本文其次通过处理深圳出租车行驶轨迹数据，建立了核密度估计模型和 Top-k 模型给出了充电桩的排布方案。构建了从地图数据，行驶数据集成在一起的解决方案。

第二章 文献综述

因为研究的问题主要如何利用已有的车辆行驶轨迹数据给出电动车充电站的布置，文献综述部分主要包含了两方面的研究：一是电动车充电站的布置，二是轨迹数据挖掘。虽然很多传统的研究并不是数据驱动的，但这些研究都能够给予我们很多启发，为数据驱动的办法提供了许多参考。

2.1 电动车充电站的布置

充电站的建造对整个经济体系都至关重要。作为一个城市的基础设施，他的建造的位置，容量大小都影响着总体的城市规划，包括电力网络，环境问题，便利程度等等。接下来主要讨论充电站的位置选取和大小。

2.1.1 传统做法

在过去的研究中通常将这个问题化为优化问题。比如有使用平衡模型框架求解的 [1]。以葡萄牙的里斯本为例 [2]，有研究利用了和传统加油站站一样的方式将问题划归为最大覆盖模型解出了所需的充电站模型和大小。除此之外，也有利用数据包络分析求解的办法。[3] 数据包络分析是一个对多投入或多产出的多个决策单元的效率评价方法，以最大化分发系统管理者的利益选择决定充电站的位置和大小。也有研究 [4] 以德国为例，提出使用动态空间模型解得充电站的需求，并且给出了直到 2020 年德国充电基建的发展展望。也有学者以北京为例 [5] 提出不仅要从估计电动车密度的角度出发，也要考虑不同的种类的充电站，以及对电网的影响。也有以司机为研究目标的充电站排布方案 [6]。它以司机的生活轨迹为研究目标，研究了电动车市场，从而寻找较优的初始充电站应该排布的位置。也有研究 [7] 以历史车流量数据为验证，为韩国的高速路提出了一个多阶段的优化布置方案。当给定电池容量的基础上求解最优的充电站位置的研究 [8]。也有建立了以到达充电站最小距离为目标的数学模型，并且利用禁忌搜索策略寻找最优解的研究 [9]。若以可以完成来回路线的人数为目标，也可以建立相应的整数规划模型 [10]。若以电动车总耗费角度出发，也可以建立相应的线性整数规划模型 [11]。

2.1.2 多因素综合决策

因为要考虑到城市电网的结构，有研究以能量平衡的角度出发设计充电站位置和大小 [12]。也有同时考虑了交通和电网供给两个因素的模型 [13]。为响应国家可持续发展的号召，充电站的排布也要可持续。从经济，社会，环境等方面均有考虑。有研究 [14] 利用了 Multiple-Criteria Decision Making (MCDM) 来决定位置。具体来说利用了 Technique for Order of Preference by Similarity to Ideal

Solution (TOPSIS) 的技术手段。有一些评判标准并不是什么准确而且有的有缺失，所以就利用了模糊集的办法。也有研究同时考虑了周边环境因素和服务半径的工作 [15]。具体的它采用了两步筛选法，它也提出了一种新的改进的原始-对偶内点算法。为了增加太阳能板的穿透能力以及减少车辆荷载的带来的边际影响，[16] 建立了多元随机模型，以最小能量耗损和电压波动为目标，利用粒子群优化算法求解。有工作以 [17] 以电池充电行为和用户充电和旅行行为为考量，利用概率分布函数求解充电站的位置。也有划归为整数规划的模型求解充电站位置的研究 [18]。除了最小化到达充电站的费用，同时也考虑了停车场，当地工作，人口密度等影响因素。还有为快速充电站设计位置的考虑多个因素的模型 [19]。考量了诸如充电站费用，电动车能量损失，电网能量损失城市道路等因素。

2.1.3 大数据驱动的办法

固然套用类似于加油站排布方案类似的做法是可行的，但是数据驱动的办法有助于我们去解决充电站排布的新问题。[20] 有相关的研究工作就指出可以利用轨迹数据进行决策，而不需要许多方案中要求的出发地和目的地作为输入。因为轨迹数据比较容易得到。解决方案分为三步，子轨迹的分离，以最小的寻找时间为目标决定充电站的位置，最后利用排队论的办法给每一个充电站分配充电桩。在日本有学者专门针对日本独特的按需供给巴士系统设计充电站的排布 [21]。这项研究分析了公交系统的子行程，提出了如何安排公交行程的办法。利用了出租车的轨迹数据估计需求，以增加顾客接受率和减少旅行距离为目标做了优化得到充电站的位置。也有研究 [22] 以加油站为基础，认为充电站应该建立在加油站附近，以出租车数据估计需求寻找优解。还有研究 [23] 以行为事件为基础，利用出行数据作为支撑，讨论了对不同出行需求的人口使用电动车的便利程度，从而寻找较优的充电站布置方案。

2.2 轨迹数据挖掘

现如今许多车上都配备了 GPS(全球定位系统) 设备，车辆可以依靠这些设备定位，方便出行。与此同时，这些设备也可以收集得到海量的车辆行驶轨迹的数据。[24] 这些数据将可以被用来开发新的应用或者提高改善现有的机遇传统方法的应用。比如：道路搜寻，目的地或者位置预测，某种群体运动行为的分析，以及其他的市政服务。这些应用将极大改善人类的生活。

然而，对于轨迹数据的管理，处理和挖掘仍然是非常具有挑战性的。比如：数据积累很快，难以管理；数据不好定义相似性，难以进行进一步的挖掘；处理那么大量的数据将十分的耗时耗力。总的来说轨迹数据的挖掘的框架分为三个层次：1. 数据的收集，2. 轨迹数据挖掘的技术，3. 应用。其中轨迹数据挖掘

技术中主要分为五个组成部分：**预处理**，**数据管理**，**数据查询**，**挖掘任务**和**隐私保护**。接下来以比较重要的数据挖掘技术和应用作为主要介绍部分。

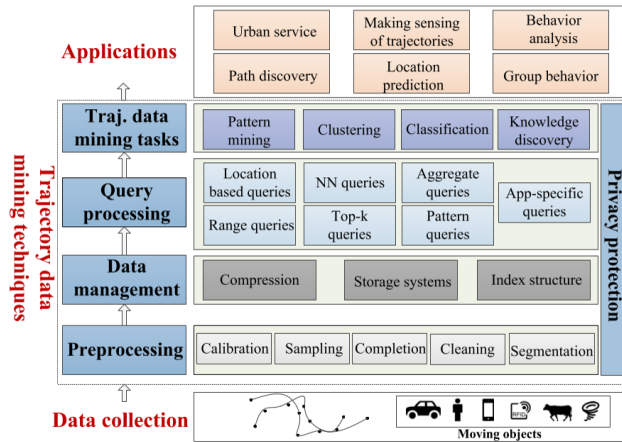


图 2.1 Trajectory Data Framework[24]

2.2.1 数据预处理

在预处理中，常用的有五个操作。**数据清理**：这一步主要是为了剔除数据中的异常点。异常点非常影响最后的挖掘质量 [25]。主要是通过筛选剔除不可能的地点和超出相应的限制（比如：最大时速，不可能达到的条件）的记录。[26] 第二个操作是**分段**，每条轨迹可以分为子轨迹 [27]。子轨迹对应于不同的行驶状态，如不同的行驶路段。良好的分段也便于解读。[28] 优秀的储存系统也将得益于分段。[29][30] 第三种操作是**数据补完**。由于数据是通过很低的采样率得到的，是不能完全反应行驶轨迹的。有不少研究致力于补完行驶轨迹数据。[31][32][33][34][35] 第四种操作是**矫正**，由于数据可能是由不同的设备采样的，然而数据是离散表示的，采样率将可能不同。这将导致记录之间的比较什么困难。[36][37] 致力于将异源数据合并统一。最后一个操作是**抽样**，由于轨迹数据的数据量十分之大，操作将变得极为耗时和复杂。相关研究工作 [38][39] 旨在抽取其中最能够表现轨迹的数据点简化操作。

2.2.2 数据管理

数据管理是轨迹数据挖掘的基本问题，良好的数据管理有助于储存和处理大规模数据集。主要有两方面的研究：**数据压缩**以及 **储存系统和指标结构**。

数据压缩 储存和转化大量的历史行驶轨迹数据是非常麻烦。因为很多轨迹中的数据点是非常冗余的，有许多的轨迹压缩算法被提出来用以减少存储的需求，同时也减少了通讯的负担。[40–42]。压缩数据的代价通常是数据准确率的下降，一般来说压缩算法都会在准确率和压缩率之间做出平衡。越高的压缩率一般会

导致越差的数据质量。传统的做法是线段产生法和 Delta 压缩法。Trajic[40] 构造了预判器从当前轨迹预判下一个轨迹点。因为实际位置和估计位置存在误差, 所以利用了残差编码的方式补偿了这个差别。这个算法在允许一定误差的情况下, 达到了很好的压缩率。PRESS[41] 提出了将空间表达和时间表达分开处理压缩的想法。在分离之后, 分别提出了时间和空间上的压缩算法。因为是可以分别处理的, 所以在并行下算法能有很高的效率。时空查询也可以在不完全解压缩的状态下进行。但是这也是有损压缩。[42] 提出了一种在线的有误差上限压缩系统。这个方法利用了以初始点为中心的虚拟坐标系统, 利用凸包划定了点的边界。另外一些研究认为应该简化原始的数据, 这样的化简算法也可以认为是一种有损压缩。[43, 44]

储存系统和指标结构 另外一个关键的问题就是如何对海量的数据进行储存。[29, 30, 45–48]。比如 SharkDB[29, 30], 这是一个轨迹数据储存系统。它能够快速的处理数据查询功能。为了应用以列为导向的储存, 轨迹首先做了分段切片。这些轨迹片段被进一步压缩和储存以支持轨迹数挖掘。Trajstore[45] 是一个动态存储系统, 它可以快速地选取特定区域内的所有轨迹数据。这个系统也是通过首先将数据切片分段, 然后根据时空的区域性, 按区域储存, 从而满足区域查询的要求。TrajTree[46] 是一种指标结构, 他被用来管理轨迹数据。它适用于类似 k-NN 的查询。也有研究工作 [47] 提出了一种适用于数据流的指标结构系统。这个工作也尝试达到一种较优的时空查询代价。另外一个相关工作是 [48], 它提出了一种参数空间的指标办法。它利用了多项式近似去为轨迹数据的线段进行标号。

2.2.3 数据查询

在有数据管理系统的基础上如何进行数据的查询也是另外一个关键的问题。以下介绍轨迹数据处理中常用的几个查询。**以位置为基础的查询**是给定了需要查询的位置, 寻找数据库中与给定位置相近的数据。其中给定的一系列地理位置也分为有序和无序的差别。一个经典的应用是路线推荐: 给定初始地点和目的地以及一系列中间要经过的地点, 寻找相应的路径。有相关的研究 [49] 通过找出 k 个最好的链接了所有给定点的路径 (k-BCT) 来提供查询功能。另外一个研究工作 [50] 是根据地点的重要性来给出查询。具体来说, 一个在地图上有标识的地点会比一个没有意义的地点更加重要。查询能够给出 k 个最重要的连接的轨迹 (k-ICT)。**范围查询**对许多轨迹数据挖掘应用有很大的帮助。具体来说他可以对时空间上有一定的约束, 比如可以查询在特定时间段内的轨迹, 也可以查询在固定经纬度范围内的轨迹数据。最近有不少研究专注于在不确定轨迹数据上的范围查询问题 [51][52]。[51] 首先建立了一个概率模型去表示一个物体在某一时间可能出现的地点。其次他建立了一套很有效的指标结构去处理

一定概率范围内的查询（比如：以某 0.95 置信区间寻找某个物体可能出现的位置）。另外一个类似的工作 [52] 提供了查询哪些轨迹在一段时间内都以很高的概率出现在某一区域的办法。**最近邻查询** [53–55] 也是另外一个基础的查询。有研究 [53] 提出了在某一特定时间段查询某一轨迹的邻近轨迹的办法。另外一个研究 [55] 扩展了最近邻查询到一个更加现实的情况。因为轨迹常常不是确定的，他提供了查询概率最近邻的办法。将不确定的轨迹描述成随机过程，它提出了给定特定轨迹和特定时间区间上不同的概率最近邻查询算法。**Top-k 查询**是指对于某个轨迹查询 k 个最相似的轨迹。**KSQ** [56] 研究在不确定轨迹中的这个问题。这其中的重点是量化两个不确定轨迹间的相似性。他也提出了一种新的距离度量和相应的指标结构去支持进一步的挖掘任务。**模式查询**是查询数据库中符合某种要求的轨迹。[57] 具体来说需要从数据库中找到满足某些给定顺序限制的轨迹。这些限制也包含了范围查询，最近邻查询。最重要的问题就在于寻找合适的条件表述方式，就像正则表达式一样。**集合查询** [39] 有别于传统的方法，他的查询结果是一种累计度量。它的想法来源于传统的数据库查询，相当于一种函数功能。比如通过某一特定路段的平均速度就是如此。还有其他许多特殊的查询，比如对有语义轨迹的关键字查询 [58]，对某种特殊的活动轨迹的查询 [59]，考虑了用户给定字段的位置相关查询 [60]，考虑用户个性化权重的查询 [61]。

2.2.4 挖掘任务

模式挖掘 模式挖掘是指发现单个或者多个物体的移动规律，常见的有群组模式 [62–64]，序列模式 [65] 和周期模式 [66, 67]。序列模式主要指有一定数量的轨迹共同都存在的一段子序列。周期模式的挖掘主要是为了理解物体移动的规律，通常分为探测周期和挖掘周期运动行为两步。

聚类 聚类主要就是将相似的运动轨迹聚类成组 [68–70]。TODMIS [69] 提供了一个挖掘小群体的框架，其中不仅考虑了轨迹相关的信息，也考虑了位置的语义信息。另一工作 [70] 解决了识别移动车辆热点的办法，而实际上这就聚类问题。与密度聚类不同，它使用移动性来聚类。比如说车移动速度越快，就越不拥挤，反之亦然。

分类 利用已知标签信息的轨迹建模估计未知标签轨迹的标签。有工作 [71] 就利用了过程信息提高了判断的准确率。很多轨迹数据独特的信息也能被拿来作判断，如空间位置，空间分布，轨迹形状和过程信息。

知识发现 很多时候我们也能得益于从数据中取得的新知识。就有研究专注于发现城市不同的功能区域 [72][73]。还有研究从不同方面探测了事件的发生。

[74, 75]

2.2.5 隐私保护

如何在数据挖掘的同时保护用户的隐私也是一个研究的热点 [76–79]。其中有工作 [76] 根据后续处理的需要将原始数据投影到一个抽象的语义空间。SmartTrace[77] 则提出了一种方法去寻找与给定轨迹最相似的轨迹。它提出的分布式相似度量可以让用户不必上传他们自己的数据，从而做到了隐私保护。也有工作 [78] 开发了一种轨迹数据处理引擎，对访问有严格的限制，保护用户的隐私。从加密的角度保护隐私也有相关的工作 [79]

2.2.6 轨迹数据挖掘的应用

有非常多的轨迹数据挖掘的应用相关的研究。主要分为 6 个不同的子类。

路径发现 这是一个最为流行的应用。如何找到一条最优的道路是一个问题。不同的问题回对应与不同的最优的定义。最优的道路可能是最短的，最快的，走得最多的。[80–85] 相比最快和最短路径，最高频的路径反应了之前人的习惯，说明是最可靠的路径更贴近实际。对于不同人的驾驶偏好有人开发了个性化路径推荐的系统 [85]。其中除了对个人行车路线的推荐之外，也有利用轨迹数据建立公共交通系统的研究 [84]。

地点或者目的地的预测 位置服务 (Location Based Services, LBS)，对人们在城市里的生活提供了极大的便利。人类活动是非常有规律且可以预测的。许多基于地理位置的应用需要位置预测和目的地预测以便给目标人群发送广告或者出现建议。比如餐厅景点的推荐和导航系统的目的地推荐。这项任务与路径发现有很深的联系。如果当前路径能够匹配到数据库中的高频路径就可以通过历史数据预测到目的地。有研究 [86][87] 指出稀疏性会使问题变难，通过将路径拆分成成为子路径再组合成新路径可以扩充数据库大小以解决稀疏的问题。通过分析移动模式去预测用户下一个将要到达的位置也是常用的手段 [88]。也有研究不仅以单个个体运动为分析基础，社会群体对某一个体的影响也被纳入考虑以达到更好的效果 [89]。

运动行为分析 轨迹数据提供了很多可以分析物体运动行为的机会 [90–95]。这其中很重要的一个目标是抽象出移动物体的高级语义信息，如目标和规则。在特定情况下人类的运动模式对市政建设很有参考意义。比如有工作 [91] 研究了面对灾难的时候人们的运动行为并建立了预测模型，这对灾难预警管理，社会重建有重要的参考意义。根据历史行驶数据也可以构建判别移动物体的角色的模型 [92]。也有研究从时间角度分析运动行为并发现了很强的周期性 [93]。

群组行为分析 运动物体尤其是人类和动物在移动时都有强烈的群组效应。人类的行为很多时候就不是只受个人事物的影响，个人所属的群体所带来的影响是不可忽略的。例子有节日庆典，游行，交通拥塞，示威抗议，大规模的商业活动。也有大量的相关研究 [62–64, 96, 97]。其中有学者提出了由密集群组产生的聚集模式 [62]。[63] 在线版本的相关工作也被提出了。也有相关研究建了模拟人类个体和群体运动的机制 [96]。它提供了通常现实中不易获得的大规模移动的数据。

市政服务 另外一个最重要的应用是为市政建设提供指导，从而提升人们的生活质量 [20, 98–100]。有相关研究从轨迹数据重建道路地图 [98]。交通流量的估计也是一个非常基础的任务，在风险评估，服务质量和位置排序中都有应用。[99] 对封闭环境中行人流量的估计就可以被应用在场布置上，比如在那些地方放咨询台，洗手间商店等。类似的应用还有提供停车信息的服务 [100]。

可解释轨迹 原始的轨迹数据是一系列的地理位置和时间，并不包含任何的语义信息。比如，工作中，购物中，看病中，等等标签都是无法直接获得。大量的研究尝试从轨迹数据中推断语义信息 [101–104]。STMaker[101] 就是一个这样的系统，他先通过根据物体的行为对轨迹的分段，然后对每一段进行总结并得到一个文字语义描述。TOPTRAC[102] 是另外一个通过寻找轨迹的潜在主题获得语义的系统。也有基于聚类算法的系统 [104]

第三章 数据驱动的充电桩布置办法

3.1 问题定义

现在的电动车充满电大约需要几个小时。充满电的时候不同电动车可以行驶的距离是不同的。然而在我们的数据集中电动车都是出租车。所以型号是一样的，所以不考虑每个出租车型号带来的个体差异，如行驶距离和电池容量充电时间等。这种电动车充满电通常可以行驶 200 公里，充满电的时间大约需要 1.5 到 2 个小时。由于可再生能源的迅速发展，越来越多的电动车出现在了城市的街道上。在深圳截至 2013 年 11 月就有接近一千辆电动出租车。截至获得数据的时间（2014 年 4 月 14 日），总共只有 26 个电动车充电站在深圳。更多的充电站需要被建立，以满足不断增长的需求。接下来通过获得的电动车出行历史轨迹数据，我们探索了如何寻找最优的新的充电站布置位置。

定义 1 轨迹 一条轨迹是指通过车辆上的 GPS 设备收集得到的一系列带有时间的连续地点。每一条记录包含经度，纬度，时间和 ID。我们把车辆行驶的轨迹首先拆分成子轨迹，便于后续分析。这也是在轨迹数据挖掘中非常常用的技术手段。可以将子轨迹区分为，旅行中，寻找充电站中，和充电中。

定义 2 充电子轨迹 充电子轨迹说明汽车停留在充电站不动并且在充电。

定义 3 寻找事件 寻找子轨迹说明汽车在落下最后一名乘客后的情况下，由于需要充电决定出发前去充电站而不是接下一名乘客的一条记录。

定义 4 寻找子轨迹 寻找子轨迹说明汽车在触发寻找事件后出发前去充电站的轨迹。

定义 5 旅行子轨迹 旅行子轨迹是代表汽车除去寻找子轨迹和充电子轨迹之外的轨迹。

如何标定三种不同的子轨迹将在解决方案一节叙述。图3.1就展现了两辆车的三种不同行驶子轨迹。

由于是城市道路度量两个地点的距离最好是使用曼哈顿距离而不是直接的欧式距离。我们的目标就是寻找从寻找事件到充电站期望意义下最短的配置方案。

问题描述 给定了一个已有的 L 个充电站的集合，包括他们的地点。现在要新建 K 个新的充电站。给出从寻找事件到充电站期望意义最短路期望最小的方案。

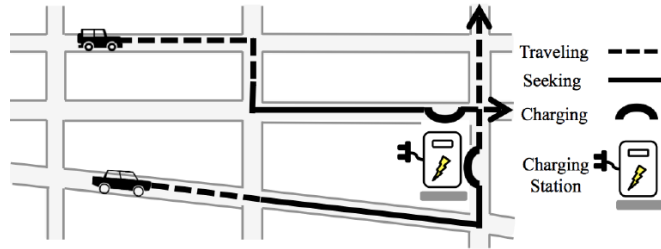


图 3.1 三种子轨迹 [20]

3.2 数据描述

本项目使用了三个数据源: 1. 深圳市城市道路地图, 2. 电动车行驶轨迹数据, 3. 现有的充电站位置数据。为了统一性, 这些数据都为同一时间采样得到。

深圳市城市道路地图 我们首先使用了 Google Geocoding API 获得了深圳市的矩形边框。东北角为 (22.70385, 114.33991), 西南角为 (22.447203, 113.769263)。通过确定了边框, 我们使用了 OpenStreetMap 获得了深圳市的道路数据, 其中包含了各种大小的路段和重要建筑的位置标识。然后使用了 PostgreSQL 和 PostGIS 管理地图数据。利用了 TileMill 实现了可视化。

电动车行驶轨迹数据 获得了大小为 5 个 G 的 2014 年 4 月 14 日的电动车行驶轨迹数据。每一条记录中包含了我们所需的重要信息, 有车辆的编号, 时间, 经度, 维度和载客的信息。

现有的充电站位置数据 截至 2014 年 4 月深圳市一共有 26 个充电站。图3.2结合了地图数据标出了当时充电站的位置。我们可以发现充电站的位置和主要交通道路的关系还是非常密切的。

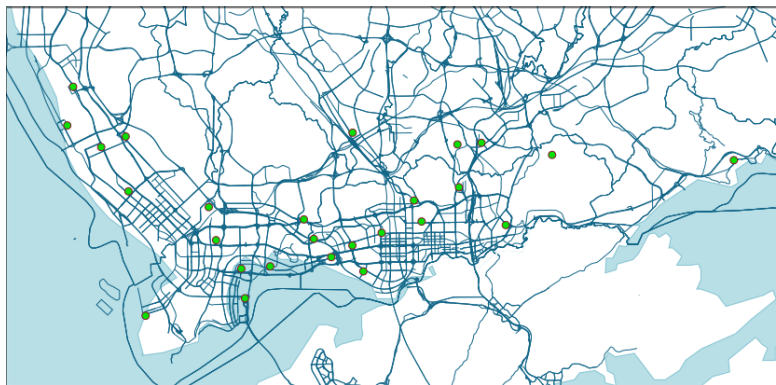


图 3.2 充电站位置和道路图, 其中绿色的为充电站

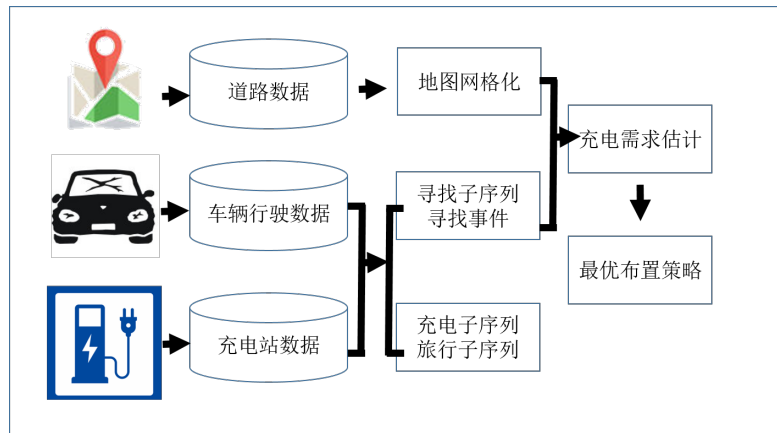


图 3.3 解决方案框架

3.3 解决方案

图3.3展示了解决方案的框架。它以三个数据库为输入，包括道路地图，充电站位置，轨迹数据。包含了四个重要的步骤，分别是 1. 地图分块，2. 子轨迹抽取，3. 需求核密度估计，4. 最优分配方案。

3.3.1 地图分块

首先道路地图被分成 n 个等同大小的区块 g_i ，这里我取的是以经纬度 0.01 为长度分块。由于充电站的布置不仅仅依赖于充电需求，还要考虑周边的地理环境和其他市政设施。所以我们仅以这种小的区块为考虑的最小元素。只要确定了在这个区块内建立充电站，就可以再具体到这个小区块内寻找合适的建立地点就行了。当然这种办法的区间区分长度可以非常灵活的确定。这也被许多其他的工作所采用。相比 Voronoi 分块，和以道路为基础的方法，这种方法简单直观。图3.4就展示了分块后的结果。

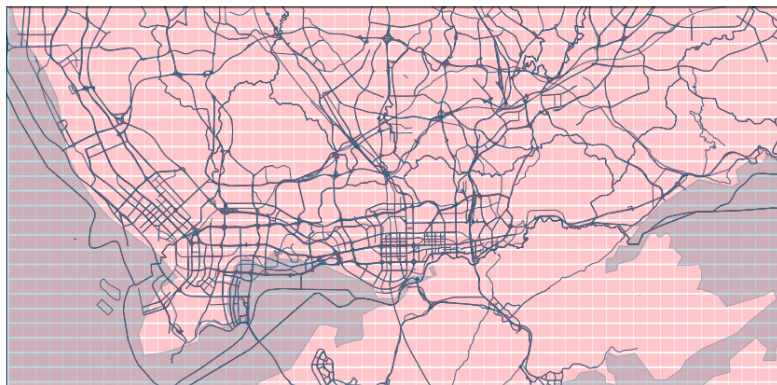


图 3.4 道路地图分块

3.3.2 子轨迹抽取

一辆车的行驶轨迹是由三种子轨迹构成的。我们首先抽取充电子轨迹，因为这是最容易处理的。其次根据充电子轨迹，我们可以得到寻找子轨迹和寻找事件。

抽取充电子轨迹 在电动出租车充电时间在一个范围内（30 分钟到 150 分钟不等）的事实，我们可以根据已有的充电站数据去探测车辆是否在充电。具体来说，如果有一段子轨迹一直处在同一位置，而且在充电站的附近，当停留时间在固定区间内，我们就可以认为这辆电动车在充电。这一系列的 GPS 记录就可以标定为充电子轨迹。

抽取寻找子轨迹和旅行子轨迹 在获得充电子轨迹之后，我们基于如下事实抽取寻找子轨迹：1. 寻找子轨迹通常在充电子轨迹之前；2. 充电子轨迹以落下最后一名乘客开始计算直到找到充电站为止。落下最后一名乘客的事件就可以认为是触发了寻找事件。在找到所有的寻找子轨迹之后，剩下的没有标定的路径就认为是旅行子轨迹。

寻找事件可视化 如上子轨迹抽取得到的最重要的数据就是寻找事件。它为接下来的需求核密度估计提供了最直接和重要的参考。若将寻找事件在不同地图区块发生的数据在地图中显示出来可得图3.5。不同大小代表了不同量的需求。

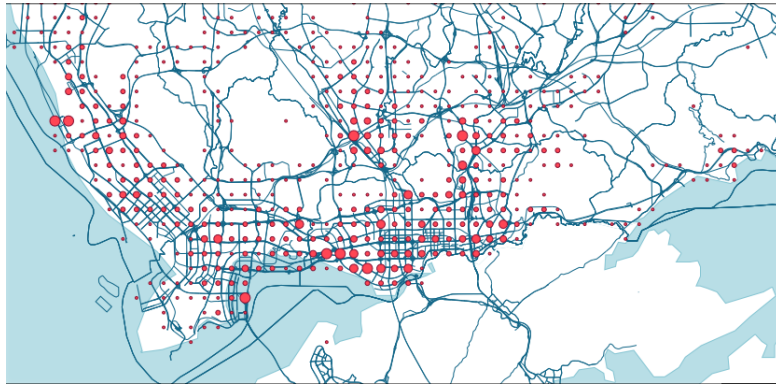


图 3.5 寻找事件可视化

Top-k 模型 当获得了每一个区块的寻找事件数量以后，选取其中最多的 k 个建立新的充电站不失为一个很好的选择。

3.3.3 需求核密度估计

为了建立进一步的模型，我们需要估计充电需求的分布。由于寻找事件的不确定性，可以假设寻找事件是在地图上的一个分布。因为每一次观测到的寻

找事件只是从这一个分布中获得的一个样例，由于从分布中获得的实例数量很多，可以利用这些实例对原来的分布进行估计。在这里我使用了高斯核密度估计的办法。得到的结果如图所示图3.6。具体来说，对于每一个地图区块 g_i ，所估计的密度为 $P_i = \sum_{j \in \text{events}} \exp(-\frac{d_{ij}^2}{2})$ ，其中 $\exp(-\frac{d_{ij}^2}{2})$ 为高斯核， d_{ij} 为区块 g_i 到事件 j 欧式距离。此处略去考虑归一化常数，因为不影响实验结果。

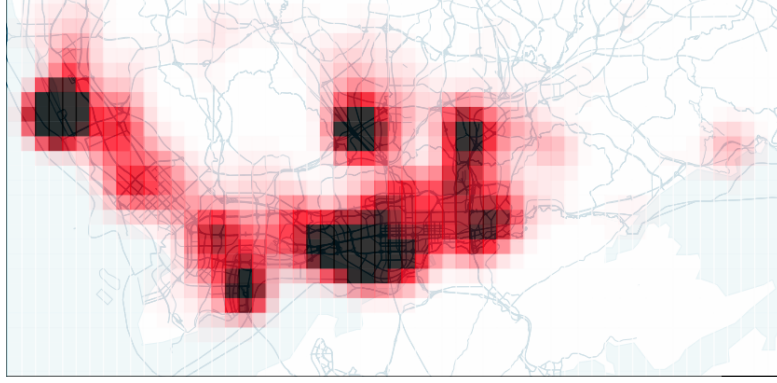


图 3.6 需求核密度估计

3.3.4 最短距离期望解

有了需求的概率分布剩下的就可以将这个问题划归数学上的整数规划的问题。给出了已经被分块好的地图 $G = g_i, (1 \leq i \leq n_0)$ ，其中 $G_L \subseteq G$ 是所有已经建立了充电站的 L 个区块的集合。由于是城市道路采用曼哈顿距离读估计两个区块 g_i, g_j 之间的距离记作 $D = [D_{ij}]$ 。让 $y = [y_i]$ 标识是否在区块 g_i 上建立充电站，为 0 时表示不建立，为 1 时表示建立。显然有如下约束 $y_i = 1, \text{for } g_i \in G_L$ 。设需要建立新的 K 个充电站，导致如下约束 $\sum_{g_i \in G} y_i = K + L$ 。用 $X = [X_{ij}]$ 表示处在 g_i 的车是否应该去 g_j 处寻找充电站。为了优化期望意义下最短的寻找距离，于是我们可以得到如下整数规划的问题。

$$\begin{aligned}
 \min \quad & z = \sum_{g_i \in G} \sum_{g_j \in G} P_i X_{ij} D_{ij} \\
 \text{s.t.} \quad & \sum_{g_j \in G} X_{ij} = 1, \quad \forall g_i \in G \\
 & \sum_{g_i \in G} y_i = K + L, \\
 & X_{i,j} \leq y_j, \quad \forall g_i, g_j \in G \\
 & X_{i,j}, y_i \in \{0, 1\}, \quad \forall g_i, g_j \in G \\
 & y_j = 1, \quad \forall g_j \in G_L
 \end{aligned}$$

如果要直接求解上述整数规划问题是比较困难而且耗费时间的，我采用了一种贪心的算法进行搜索。由于很多地方是不可能建立充电站的，平时也不会

有很多的充电需求，于是可以先利用寻找事件数量，设定一个阈值将这部分删去。其次当需要建立的充电站十分多时搜索空间程指数上涨，于是采取一次确定几个（比如在我的实验中为 5 个）充电站的位置的贪心算法去计算充电站位置。

3.4 实验结果

下面主要从两方面分析实验的结果，一个是可视化分析，一个是从期望最短寻找距离出发。

3.4.1 可视化分析

若以新增 20 个充电站为试验标准，图3.7以及图3.8分别以 Top-k 模型和核密度估计模型计算，并且在道路地图上绘制出了 20 个新产生的充电站候选位置。其中深色部分为 26 个已经存在的充电站位置，浅绿色部分为 20 个新增的充电桩候选位置。

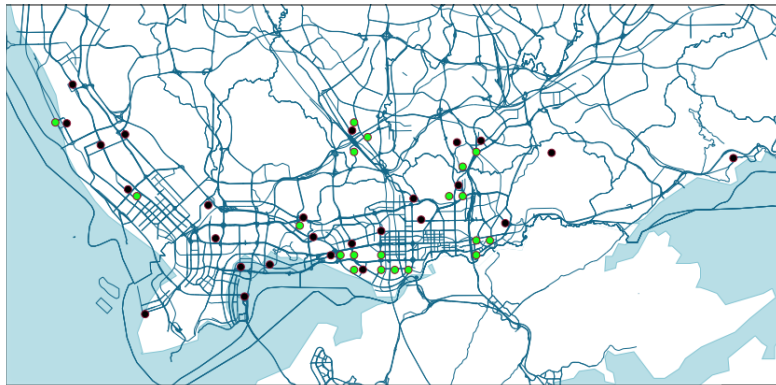


图 3.7 Top-k 模型所产生的 20 个新的充电站

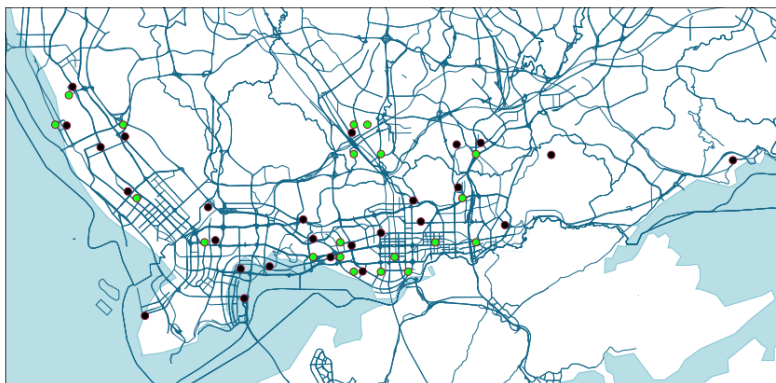


图 3.8 核密度估计模型所产生的 20 个新的充电站

从图中也可以发现由简单的 Top-k 模型产生的点虽然也能较好的符合充电需求但是存在一个问题：在福田罗湖等地用电量需求极大，所以导致这一片的

许多区块的需求量都远高于城市其他地点，Top-k 主要都建议充电站建立在此处。虽然在此处建立较多充电站确实满足了这片区域的需求，不过事实上并非真的需要建立如此密集的充电站，可能会产生冗余。如果规定了建立充电站的个数，在方便了这个区域的同时，其他区域的充电需求并没有得到很好的满足。反观由核密度估计模型所产生的 20 个新的充电站，虽然在用电多发地段安排了较多的充电站，不过并没有那么密集，充电站的建立分布更加分散一些如此受到照顾的区域更大。

3.4.2 期望最短寻找距离

若以期望最短寻找距离出发，我考察了以 Top-k 模型和核密度估计模型分别产生了从 5 到 50 不同数量的充电站对期望最短寻找距离的影响。其中期望最短寻找距离的定义为： $\sum_{g_i \in G} P_i D_{i,j(i)}$ ，其中 P_i 为区块 g_i 估计得到的出现寻找事件的期望。 $j(i)$ 为从区块 g_i 出发到区块 $g_{j(i)}$ 寻找充电站可以达到最小曼哈顿距离中 $g_{j(i)}$ 的区块编号。 $D_{i,j}$ 则是区块 g_i 出发到区块 g_j 的曼哈顿距离。

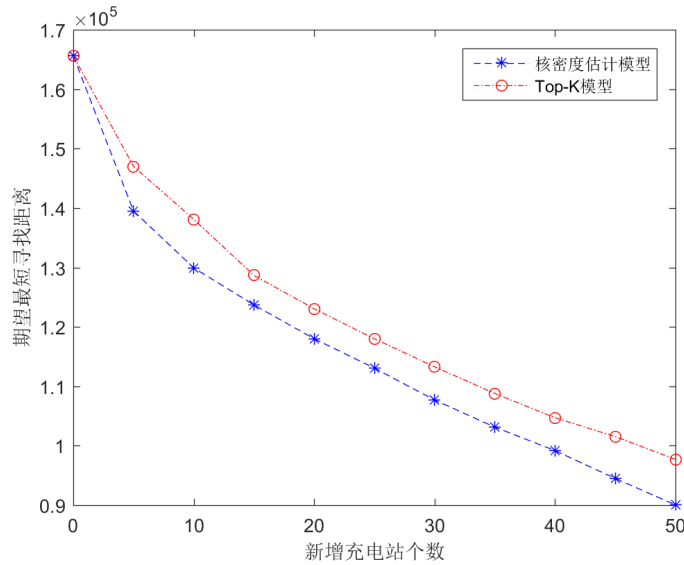


图 3.9 两个模型给出建立不同数量充电站时对应期望最短寻找距离

由于只是单位的差别，试验中并未考虑距离度量的单位和密度归一化常数。从图3.9中可以发现明显核密度估计模型的效果要明显优于 Top-k 模型。

第四章 总结

4.1 总结

为了应对不断增加的电动车充电需求，充电站的增加势在必行。本文提供了如何利用大数据建模给出优化的充电站排布方案，以及相关详尽的参考资料。

本文首先对已有的相关文献做了详尽的调研。其中主要包括两个方面：**1.充电桩排布方案 2.轨迹数据挖掘**。充电桩的排布方案，国内外已经有了非常多的探讨。有基于优化，模拟的办法。而且这个问题类似于加油站的建立排布问题，是属于传统的优化工业工程领域的重点研究问题之一。过去的研究对现在新出现的充电站排布问题有很好的参考价值。轨迹数据挖掘方面，由于轨迹数据具有数据量大，难以处理等特点，相关的研究也非常丰富。从最基本的预处理，储存，基本查询，构建上层应用都有丰富研究值得参考。详尽的文献综述不仅是对自己阅历的一个丰富，也为将来的研究提供了参考。

本文其次通过处理深圳出租车行驶轨迹数据，建立了核密度估计模型和 Top-k 模型给出了充电桩的排布方案。构建了从地图数据，行驶数据集成在一起的解决方案。具体来说分为了如下步骤：**1. 地图区块化，2. 子轨迹抽取，3. 充电需求的核密度估计，4. 优化期望最短寻找距离**。结果证明核密度估计模型要优于简单的 Top-k 模型。为将来的研究工作提供了参考。

4.2 未来工作

今后此方向的研究可以从以下几个角度出发：**1. 由于本文只考虑了距离，没有考虑行驶的速度和时间，有时候更快的速度和更短的时间也许是被人们更加看重的；2. 需求的分布只考虑了空间分布，若需要考虑行驶速度和时间的话，由于高峰期的存在，行驶时间将随时间的变化而变化，应当再考虑需求在时间上的分布。**

参考文献

- [1] He F, Wu D, Yin Y, et al. Optimal deployment of public charging stations for plug-in hybrid electric vehicles. *Transportation Research Part B Methodological*, 2013, 47(1):87–101.
- [2] Frade I, Ribeiro A, Gonçalves G, et al. Optimal Location of Charging Stations for Electric Vehicles in a Neighborhood in Lisbon, Portugal. *Transportation Research Record Journal of the Transportation Research Board*, 2011, 2252(2252):91–98.
- [3] Khalkhali K, Abapour S, Moghaddas-Tafreshi S M, et al. Application of data envelopment analysis theorem in plug-in hybrid electric vehicle charging station planning. *Iet Generation Transmission Distribution*, 2015, 9(7):666–676.
- [4] Wirges J, Linder S, Kessler A. Modelling the Development of a Regional Charging Infrastructure for Electric Vehicles in Time and Space. *European Journal of Transport Infrastructure Research*, 2012, 12(4):391–416.
- [5] Liu J. Electric vehicle charging infrastructure assignment and power grid impacts assessment in Beijing. *Energy Policy*, 2012, 51(6):544–557.
- [6] Ma T J, Zhao J J, Xiang S J, et al. An Agent-Based Training System for Optimizing the Layout of AFVs' Initial Filling Stations. *Journal of Artificial Societies Social Simulation*, 2014, 17(4):6.
- [7] Chung S H, Kwon C. Multi-period planning for electric car charging station locations: A case of Korean Expressways. *European Journal of Operational Research*, 2015, 242(2):677–687.
- [8] Wang G, Xu Z, Wen F, et al. Traffic-Constrained Multiobjective Planning of Electric-Vehicle Charging Stations. *IEEE Transactions on Power Delivery*, 2013, 28(4):2363–2372.
- [9] Xu H, Miao S, Zhang C, et al. Optimal placement of charging infrastructures for large-scale integration of pure electric vehicles into grid. *International Journal of Electrical Power Energy Systems*, 2013, 53(1):159–165.
- [10] You P S, Hsieh Y C. A hybrid heuristic approach to the problem of the location of vehicle charging stations. *Computers Industrial Engineering*, 2014, 70(1):195–204.
- [11] Baouche F, Billot R, Trigui R, et al. Efficient Allocation of Electric Vehicles Charging Stations: Optimization Model and Application to a Dense Urban Network. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 6(3):33–43.
- [12] Wang Z, Liu P, Cui J, et al. Research on Quantitative Models of Electric Vehicle Charging Stations Based on Principle of Energy Equivalence. *Mathematical Problems in Engineering*, 2014, 2013(3):831–842.
- [13] Yao W, Zhao J, Wen F, et al. A Multi-Objective Collaborative Planning Strategy for Integrated Power Distribution and Electric Vehicle Charging Systems. *Power Systems IEEE Transactions on*, 2014, 29(4):1811–1821.
- [14] Guo S, Zhao H, Yan J. Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective. *Applied Energy*, 2015, 158:390–402.
- [15] Liu Z, Wen F, Ledwich G. Optimal Planning of Electric-Vehicle Charging Stations in Distribution Systems. *Power Delivery IEEE Transactions on*, 2013, 28(1):102–110.
- [16] Pashajavid E, Golkar M A. Optimal placement and sizing of plug in electric vehicles charging stations within distribution networks with high penetration of photovoltaic panels. *Journal of Renewable Sustainable Energy*, 2013, 5(5):1–1948.
- [17] Lee Y G, Kim H S, Kho S Y, et al. User Equilibrium-Based Location Model of Rapid Charging Stations for Electric Vehicles with Batteries That Have Different States of Charge. *Proceedings of Transportation Research Board 93rd Annual Meeting*, 2014.
- [18] Chen T D, Kockelman K M, Khan M. Locating Electric Vehicle Charging Stations: Parking-Based Assignment Method for Seattle, Washington. 2013.
- [19] Sadeghi-Barzani P, Rajabi-Ghahnavieh A, Kazemi-Karegar H. Optimal fast charging station placing and sizing. *Applied Energy*, 2014, 125(2):289–299.
- [20] Li Y, Luo J, Chow C Y, et al. Growing the charging station network for electric vehicles with trajectory data analytics. *Proceedings of IEEE International Conference on Data Engineering*, 2015. 1376–1387.
- [21] Kameda H, Mukai N. Optimization of charging station placement by using taxi probe data for on-demand electrical bus system. *Proceedings of Knowledge-Based and Intelligent Information and Engineering Systems - International Conference, Kes 2011, Kaiserslautern, Germany, September 12-14, 2011, Proceedings*, 2011. 606–615.

- [22] Cai H, Jia X, Chiu A S F, et al. Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet. *Transportation Research Part D Transport Environment*, 2014, 33:39–46.
- [23] Dong J, Liu C, Lin Z. Charging infrastructure planning for promoting battery electric vehicles: An activity-based approach using multiday travel data. *Transportation Research Part C Emerging Technologies*, 2014, 38(1):44–55.
- [24] Feng Z, Zhu Y. A Survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access*, 2016, 4:2056–2067.
- [25] Ge Y, Xiong H, Zhou Z H, et al. Top-Eye: top-k evolving trajectory outlier detection. *Proceedings of ACM Conference on Information and Knowledge Management, CIKM 2010*, Toronto, Ontario, Canada, October, 2010. 1733–1736.
- [26] Fazzinga B, Flesca S, Furfaro F, et al. Cleaning trajectory data of RFID-monitored objects through conditioning under integrity constraints. *Proceedings of EDBT*, 2014. 379–390.
- [27] Pelekis N, Kopanakis I, Panagiotakis C, et al. Unsupervised Trajectory Sampling. 2010. 17–33.
- [28] Su H, Zheng K, Zeng K, et al. Making sense of trajectory data: A partition-and-summarization approach. *Proceedings of IEEE International Conference on Data Engineering*, 2015. 963–974.
- [29] Wang H, Zheng K, Xu J, et al. SharkDB: An In-Memory Column-Oriented Trajectory Storage. *Proceedings of ACM International Conference on Conference on Information and Knowledge Management*, 2014. 1409–1418.
- [30] Wang H, Zheng K, Zhou X, et al. SharkDB: An In-Memory Storage System for Massive Trajectory Data. *Proceedings of ACM SIGMOD International Conference on Management of Data*, 2015. 1099–1104.
- [31] Zheng K, Zheng Y, Xie X, et al. Reducing Uncertainty of Low-Sampling-Rate Trajectories. *Proceedings of IEEE International Conference on Data Engineering*, 2012. 1144–1155.
- [32] Chiang M F, Lin Y H, Peng W C, et al. Inferring distant-time location in low-sampling-rate trajectories. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013. 1454–1457.
- [33] Li M, Ahmed A, Smola A J. Inferring Movement Trajectories from GPS Snippets. *Proceedings of Eighth ACM International Conference on Web Search and Data Mining*, 2015. 325–334.
- [34] Banerjee P, Ranu S, Raghavan S. Inferring Uncertain Trajectories from Partial Observations. *Proceedings of IEEE International Conference on Data Mining*, 2014. 30–39.
- [35] Larusso N D, Singh A. Efficient tracking and querying for coordinated uncertain mobile objects. *Proceedings of Data Engineering (ICDE), 2013 IEEE 29th International Conference on*. IEEE, 2013. 182–193.
- [36] Su H, Zheng K, Wang H, et al. Calibrating trajectory data for similarity-based analysis. *Proceedings of ACM SIGMOD International Conference on Management of Data*, 2013. 833–844.
- [37] Su H, Zheng K, Huang J, et al. Calibrating trajectory data for spatio-temporal similarity analysis. *The VLDB Journal*, 2015, 24(1):93–116.
- [38] Panagiotakis C, Pelekis N, Kopanakis I, et al. Segmentation and Sampling of Moving Object Trajectories Based on Representativeness. *IEEE Transactions on Knowledge Data Engineering*, 2012, 24(7):1328–1343.
- [39] Li Y, Chow C Y, Deng K, et al. Sampling Big Trajectory Data. *Proceedings of The ACM International*, 2015. 941–950.
- [40] Nibali A, He Z. Trajic: An Effective Compression System for Trajectory Data. *IEEE Transactions on Knowledge Data Engineering*, 2015, 27(11):3138–3151.
- [41] Song R, Sun W, Zheng B, et al. PRESS: a novel framework of trajectory compression in road networks. *Proceedings of the Vldb Endowment*, 2014, 7(9):661–672.
- [42] Liu J, Zhao K, Sommer P, et al. Bounded Quadrant System: Error-bounded Trajectory Compression on the Go. 2014. 987–998.
- [43] Long C, Wong R C W, Jagadish H. Direction-preserving trajectory simplification. *Proceedings of the VLDB Endowment*, 2013, 6(10):949–960.
- [44] Long C, Wong R C W, Jagadish H. Trajectory simplification: on minimizing the direction-based error. *Proceedings of the VLDB Endowment*, 2014, 8(1):49–60.

- [45] Cudre-Mauroux P, Wu E, Madden S. Trajstore: An adaptive storage system for very large trajectory data sets. *Proceedings of Data Engineering (ICDE), 2010 IEEE 26th International Conference on*. IEEE, 2010. 109–120.
- [46] Ranu S, Deepak P, Telang A D, et al. Indexing and matching trajectories under inconsistent sampling rates. *Proceedings of Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 2015. 999–1010.
- [47] Sandu Popa I, Zeitouni K, Oria V, et al. Indexing in-network trajectory flows. *The VLDB Journal—The International Journal on Very Large Data Bases*, 2011, 20(5):643–669.
- [48] Ni J, Ravishankar C V. Indexing spatio-temporal trajectories with efficient polynomial approximations. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(5).
- [49] Chen Z, Shen H T, Zhou X, et al. Searching trajectories by locations: an efficiency study. *Proceedings of Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, 2010. 255–266.
- [50] Yan D, Cheng J, Zhao Z, et al. Efficient location-based search of trajectories with location importance. *Knowledge and Information Systems*, 2015, 45(1):215–245.
- [51] Zheng K, Trajcevski G, Zhou X, et al. Probabilistic range queries for uncertain trajectories on road networks. *Proceedings of Proceedings of the 14th International Conference on Extending Database Technology*. ACM, 2011. 283–294.
- [52] Zhan L, Zhang Y, Zhang W, et al. Range search on uncertain trajectories. *Proceedings of Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015. 921–930.
- [53] Güting R H, Behr T, Xu J. Efficient k-nearest neighbor search on moving object trajectories. *The VLDB Journal—The International Journal on Very Large Data Bases*, 2010, 19(5):687–714.
- [54] Kim G J, Park Y H, Cho W H. Efficient Nearest Neighbor Search on Moving Object Trajectories. *Journal of the Korea Institute of Information and Communication Engineering*, 2014, 18(12):2919–2925.
- [55] Niedermayer J, Züfle A, Emrich T, et al. Probabilistic nearest neighbor queries on uncertain moving object trajectories. *Proceedings of the VLDB Endowment*, 2013, 7(3):205–216.
- [56] Ma C, Lu H, Shou L, et al. KSQ: Top-k similarity query on uncertain trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(9):2049–2062.
- [57] Vieira M R, Bakalov P, Tsotras V J. Querying trajectories using flexible patterns. *Proceedings of Proceedings of the 13th International Conference on Extending Database Technology*. ACM, 2010. 406–417.
- [58] Zheng B, Yuan N J, Zheng K, et al. Approximate keyword search in semantic trajectory database. *Proceedings of Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 2015. 975–986.
- [59] Zheng K, Shang S, Yuan N J, et al. Towards efficient search for activity trajectories. *Proceedings of Data Engineering (ICDE), 2013 IEEE 29th International Conference on*. IEEE, 2013. 230–241.
- [60] Shang S, Ding R, Yuan B, et al. User oriented trajectory search for trip recommendation. *Proceedings of Proceedings of the 15th International Conference on Extending Database Technology*. ACM, 2012. 156–167.
- [61] Shang S, Ding R, Zheng K, et al. Personalized trajectory matching in spatial networks. *The VLDB Journal*, 2014, 23(3):449–468.
- [62] Zheng Y, Yuan N J, Zheng K, et al. On discovery of gathering patterns from trajectories. *Proceedings of IEEE International Conference on Data Engineering*, 2013. 242–253.
- [63] Zheng K, Zheng Y, Yuan N J, et al. Online Discovery of Gathering Patterns over Trajectories. *IEEE Transactions on Knowledge Data Engineering*, 2014, 26(8):1974–1988.
- [64] Li X, Ceikute V, Jensen C S, et al. Effective Online Group Discovery in Trajectory Databases. *Knowledge Data Engineering IEEE Transactions on*, 2013, 25(12):2752–2766.
- [65] Zhang C, Han J, Shou L, et al. Splitter: mining fine-grained sequential patterns in semantic trajectories. *Proceedings of the Vldb Endowment*, 2014, 7(9):769–780.
- [66] Li Z, Ding B, Han J, et al. Mining periodic behaviors for moving objects. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, Dc, Usa, July, 2010. 1099–1108.
- [67] Li Z, Han J, Ding B, et al. Mining periodic behaviors of object movements for animal and biological sustainability studies. *Data Mining Knowledge Discovery*, 2012, 24(2):355–386.

- [68] Vries G D, Someren M V. Clustering Vessel Trajectories with Alignment Kernels under Trajectory Compression. Springer Berlin Heidelberg, 2010: 296–311.
- [69] Liu S, Wang S, Jayarajah K, et al. TODMIS: mining communities from trajectories. Proceedings of ACM International Conference on Conference on Information Knowledge Management, 2013. 2109–2118.
- [70] Liu S, Liu Y, Ni L M, et al. Towards mobility-based clustering. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, Dc, Usa, July, 2010. 919–928.
- [71] Patel D, Sheng C, Hsu W, et al. Incorporating duration information for trajectory classification. Proceedings of Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE, 2012. 1132–1143.
- [72] Yuan N J, Zheng Y, Xie X, et al. Discovering Urban Functional Zones Using Latent Activity Trajectories. Knowledge Data Engineering IEEE Transactions on, 2015, 27(3):712–725.
- [73] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012. 186–194.
- [74] Wang X, Li G, Jiang G, et al. Semantic trajectory-based event detection and event pattern mining. Knowledge Information Systems, 2013, 37(2):305–329.
- [75] Au T S, Duan R, Kim H, et al. Spatiotemporal Event Detection in Mobility Network. 2010. 28–37.
- [76] Andrienko N, Andrienko G, Fuchs G, et al. Visual Analytics Methodology for Scalable and Privacy-Respectful Discovery of Place Semantics from Episodic Mobility Data. Springer International Publishing, 2015: 254–258.
- [77] Costa C, Laoudias C, Zeinalipouryazti D, et al. SmartTrace: Finding similar trajectories in smartphone networks without disclosing the traces. Proceedings of IEEE International Conference on Data Engineering, 2011. 1288–1291.
- [78] Pelekis N, Gkoulalas-Divanis A, Voudas M, et al. Privacy-aware querying over sensitive trajectory data. Proceedings of ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October, 2011. 895–904.
- [79] Kong L, He L, Liu X Y, et al. Privacy-Preserving Compressive Sensing for Crowdsensing Based Trajectory Recovery. Proceedings of IEEE International Conference on Distributed Computing Systems, 2015. 31–40.
- [80] Lu H C, Lee W C, Tseng V S. Mining fastest path from trajectories with multiple destinations in road networks. Knowledge and Information Systems, 2011, 29(1):25–53.
- [81] Chen Z, Shen H T, Zhou X. Discovering popular routes from trajectories. Proceedings of IEEE International Conference on Data Engineering, 2011. 900–911.
- [82] Wei L Y, Zheng Y, Peng W C. Constructing popular routes from uncertain trajectories. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012. 195–203.
- [83] Luo W, Tan H, Chen L, et al. Finding time period-based most frequent path in big trajectory data. Proceedings of ACM SIGMOD International Conference on Management of Data, 2013. 713–724.
- [84] Liu Y, Liu C, Yuan N J, et al. Exploiting Heterogeneous Human Mobility Patterns for Intelligent Bus Routing. Proceedings of IEEE International Conference on Data Mining, 2014. 360–369.
- [85] Dai J, Yang B, Guo C, et al. Personalized route recommendation using big trajectory data. Proceedings of IEEE International Conference on Data Engineering, 2015. 543–554.
- [86] Xue A Y, Zhang R, Zheng Y, et al. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. Proceedings of IEEE International Conference on Data Engineering, 2013. 254–265.
- [87] Xue A Y, Zhang R, Zheng Y, et al. DesTeller: a system for destination prediction based on trajectories with privacy protection. 2013, 6(12).
- [88] Noulas A, Scellato S, Lathia N, et al. Mining User Mobility Features for Next Place Prediction in Location-Based Services. Proceedings of IEEE International Conference on Data Mining, 2012. 1038–1043.
- [89] Wang Y, Yuan N J, Lian D, et al. Regularity and Conformity: Location Prediction Using Heterogeneous Mobility Data. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015. 1275–1284.

- [90] Renso C, Baglioni M, Macedo J A F D, et al. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge Information Systems*, 2013, 37(2):331–362.
- [91] Song X, Zhang Q, Sekimoto Y, et al. Prediction of human emergency behavior and their mobility following large-scale disaster. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014. 5–14.
- [92] Ando S, Suzuki E. Role-Behavior Analysis from Trajectory Data by Cross-Domain Learning. *Proceedings of IEEE International Conference on Data Mining*, 2011. 21–30.
- [93] Gao H, Tang J, Hu X, et al. Modeling temporal effects of human mobile behavior on location-based social networks. *ACM*, 2013: 1673–1678.
- [94] Liu S, Qu Q, Wang S. Rationality Analytics from Trajectories. *Acm Transactions on Knowledge Discovery from Data*, 2015, 10(1):1–22.
- [95] González M C, Hidalgo C A, Barabási A L. Understanding individual human mobility patterns. *Nature*, 2008, 453(7196):779.
- [96] Gupta A, Mishra A, Vadlamudi S G, et al. A Mobility Simulation Framework Of Humans With Group Behavior Modeling. *Proceedings of IEEE International Conference on Data Mining*, 2013. 1067–1072.
- [97] Mcguire M P, Janeja V P, Gangopadhyay A. Mining trajectories of moving dynamic spatio-temporal regions in sensor datasets. *Data Mining Knowledge Discovery*, 2014, 28(4):961–1003.
- [98] Liu X, Biagioni J, Eriksson J, et al. Mining large-scale, sparse GPS traces for map inference: comparison of approaches. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012. 669–677.
- [99] Liebig T, Xu Z, May M, et al. *Pedestrian Quantity Estimation with Trajectory Patterns*. Springer Berlin Heidelberg, 2012: 629–643.
- [100] Yang B, Fantini N, Jensen C S. iPark: identifying parking spaces from trajectories. 2013..
- [101] Su H, Zheng K, Zeng K, et al. STMaker: a system to make sense of trajectory data. *Proceedings of the Vldb Endowment*, 2014, 7(13):1701–1704.
- [102] Kim Y, Han J, Yuan C. TOPTRAC: Topical Trajectory Pattern Mining. 2015. 587.
- [103] Lv M, Chen L, Chen G. Discovering personally semantic places from GPS trajectories. *Proceedings of ACM International Conference on Information and Knowledge Management*, 2012. 1552–1556.
- [104] Lu C T, Lei P R, Peng W C, et al. *A Framework of Mining Semantic Regions from Trajectories*. Springer Berlin Heidelberg, 2011.

致 谢

在中国科技大学完成本科的四年里，我所从事的学习和研究工作，都是在导师以及系里其他老师和同学的指导和帮助下进行的。在完成论文之际，请容许我对他们表达诚挚的谢意。

首先感谢导师刘恒昌的指导，教诲和关心，是他把我带到了数据挖掘的研究领域。刘老师严谨的研究态度及忘我的工作精神，都将使我受益终身。

感谢班主任郭民生老师多年的关怀。感谢刘利刚、姚新、王硕等老师，他们本科阶段的指导给我未来研究生阶段的研究工作打下了基础。

还要特别感谢孙小山等师兄的指点和照顾。

感谢科大给予我了学习第二学位的机会，感谢一路走过来的兄弟姐妹们，在最宝贵年华里，是你们伴随着我的成长。

最后，感谢我家人一贯的鼓励和支持，你们是我追求学业的坚强后盾。

姜昊茗

2017年9月26日