

# 基于网约车数据的交通数据可视化分析

万静意 黄晓丹 杨睿铭 唐 豪 晏 雁 夏晓峰

**摘要:** 基于 ChinaVis2019 挑战 2 提供的“滴滴”公司内部数据, 构建了完整的交通流量数据可视化展示系统。使用订单以及轨迹信息等对网约车的时空活动规律进行分析, 以多种可视化和文本等方式展示了赛事中心附近的交通流量演变过程和交通拥堵特征。为用户提供了多角度的可视分析, 辅助其为活动提供更好的交通疏导方案。

**关键词:** 可视化分析, 地图, 网约车数据, Echarts, 交通流量

## 简介

本文首先介绍了交通流量数据可视化展示系统的实现过程及关键技术, 然后描述了具体的可视化设计, 最后面向主办方管理与疏导的实际需求, 提出了切实可行的解决方案。

## 1 网约车时空轨迹分析流程

本文数据源分为快车专车订单时空轨迹数据以及订单起终点时间地点数据两类。本文数据分析的流程如图 1 所示。

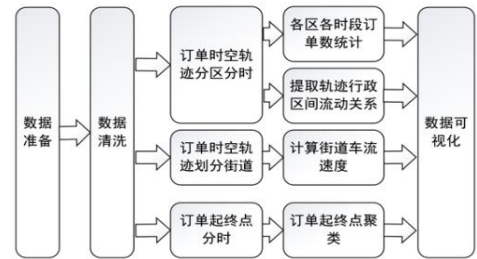


图 1 订单时空轨迹数据分析流程

在数据准备阶段, 除了使用比赛提供的数据之外, 还使用了当前天气数据和《城市道路交通管理评价指标体系》中规定的街道平均时速与道路拥堵关系来辅助问题分析。

在数据清洗阶段, 剔除了订单轨迹数据中超出当日以及时间间隔不符合 3 秒的异常点。为保证在地图中的正确显示和使用, 对轨迹经纬度数据进行了坐标转换、偏移等处理。

由于数据的时空跨度较大, 不利于详细分析数据规律和可视化展示, 所以对数据进行了分时分区处理。并结合路线起终点所在的行政区, 分析了行政区街道间的流量转移情况。

为定性判定路段拥堵情况, 引入了系统输入的交通状态参数, 即平均车速。平均车速通过随机抽取该道路上订单轨迹行驶数据与临近路段的数据加权计算得到。

对于订单起终点数据, 按照起点、终点分别聚类 and 联合聚类的方式进行聚类分析, 然后依据订单数量对聚类点进行排序, 由此可得到热门起点和到达点、热门路线的排序信息。

最后, 为了实现细粒度、海量轨迹点的时空变化可视化辅助问题分析, 使用了 Echarts、Leaflet 绘制的轨迹散点图、热力图、轨迹流动图结合多种统计图和文本的方式。

## 2 可视化设计与分析

本可视化方案针对网约车交通数据的属性及特点, 提出:

“面向流量”和“面向路线”两种车辆移动数据视图, 面向流量主要分析其流量状态。面向路线, 提出基于起点/终点单独聚类和联合聚类两种聚类方法。

### 2.1 基于时空单位的交通流量分析

基于网约车订单和轨迹数据, 从整体时空、空间、时间维度对其进行了可视化分析。

#### 2.1.1 交通流量整体时空分布规律

以动画方式在地图上展现特定时间区间内的交通情况。例如, 如图 2.1.1 所示是 18 点到 19 点的交通流量分布, 车流量主要集中在赛事中心的右上方区域, 且流量随着时间密度增大并逐渐经过赛事中心蔓延到左下方。

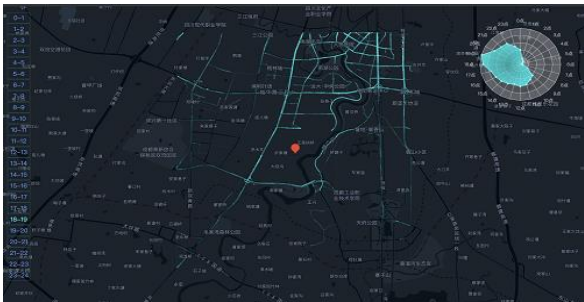


图 2.1.1 车流量总体时空分布

#### 2.1.2 交通总流量时间分布规律

赛事中心 10 公里范围内的交通总流量通过图 2.1.2 可以发现, 在一天 24 小时内, 上午段较小, 下午段到晚上流量增大, 且在 18 点到 19 点左右达到最高峰, 这也正好符合当日的实际情况, 即在赛事中心举办的活动开始时间为 19:30。



图 2.1.2 交通日流量变化

#### 2.1.3 交通流量空间分布规律

按照经纬度范围区间  $x=[103.985, 104.08499]$ ,  $y=[30.41,$

30.50999]将数据所在区域划分为了4×4方格,如图2.1.2中的黑色方格所示,管理者可根据车辆在划分区域的位置分布情况将人员和车辆引导到流量密度小的区域。

## 2.2 道路流量状态分析

为定性判定路段拥堵情况,引入了系统输入的交通状态参数,即平均车速。由于提前对数据进行了清洗筛选,因而可以抵御停车上下客、短时经纬度跳变、孤点数据异常等波动点的干扰,同时由于轨迹数据中车辆在道路交叉口停车,无法反映路段交通拥挤状态,因此对该部分数据进行了剔除。划分标准如下:

拥堵程度	平均行程车速 (km/h)
十分畅通	40 以上
畅通	30~40
轻度拥堵	20~30
拥堵	10~20
严重拥堵	10 以下

并根据此标准以5分钟为间隔将统计道路进行划分排序并实时显示,如图2.2.1(a)所示。根据道路的拥挤程度绘制了热力分布图,如图2.2.1(b)(c)所示。



图 2.2.1 道路流量状态

## 2.3 面向路线的聚类分析

交通流量演变情况与车辆的聚集状态有关,热门出发地和到达地有助于我们对全局交通状态的了解,为了获得以上信息,我们使用 Python 实现 Enhanced K-means 聚类方法,如图2.3.1所示。

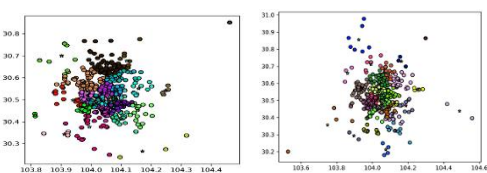


图 2.3.1 Enhanced K-means 聚类算法

为了可视化展示车流量转移特征,采用迁徙图、散点图、关系图以及玫瑰图从多个问题角度进行分析,如图2.3.2所示。



图 2.3.2 车辆轨迹时空聚类可视化

## 3 案例分析与建议

由折线图分析订单量高峰出现在18-19点,如图3(a)。查看18-19点街道订单流动关系图如图3(b),大量订单涌入协和街道、华阳镇街道以及公共街道。查看18-19点轨迹流动变化如图3(c),赛事中心周围路段轨迹点聚集,路段对应关系图中大量订单涌入的街道。说明18-19点是赛事中心举办活动的集中入场时间段。

由图3(d)(e)所示18-19点热力图观察赛事中心周围可能的拥堵点。进一步用平行坐标图分析各街道的拥堵情况,发现在街道等级为拥堵时出现了赛事中心附近的街道。结合两种图分析的拥堵点,得到乐盟国际学校南区、剑南大道南段天府第四中学、华阳客运中心、武汉路西段、云龙路为赛事中心周围的拥堵点。分析拥堵点位置发现拥堵原因包括拥堵点位于十字路口、立交桥、客运中心这些交通、人流量常汇集的地方。

根据热力图的分析可知,赛事中心在18-19点对赛事中心场馆进出口的两条路段龙马路和云龙路进行了交通限流,场馆四周存在易引发交通拥堵的路段和路段交叉口,向主办方提出的建议之一是修改赛事中心场馆附近龙马路和云龙路的交通限流方案,设置云龙路段实行车辆即停即走的方案缓解周围道路拥堵和车辆聚集情况。

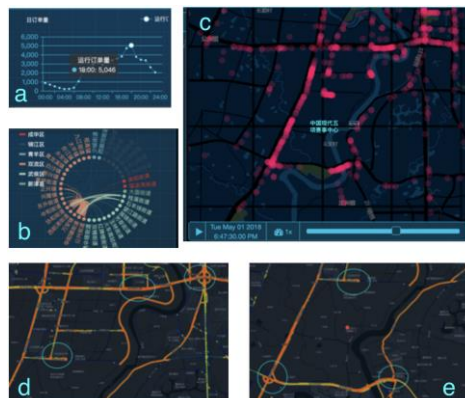


图3 (a)折线图(b)区域流动关系图(c)区域实时轨迹情况 (d)(e)18-19点热力图拥堵点分析

## 4 结束语

交通流量数据可视化展示系统,为分析场馆附近交通流量演变、分析道路交通拥堵特征、分析道路规划提供了全面的可视分析工具,能有效地帮助管理者观测交通状态,为公众提供动态、准确的交通控制和引导信息。

## 5 参考文献

- [1] Arthur, D. "k-means++: the advantages of careful seeding." Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 2007 ACM, 2007.
- [2] Jiang X R, Tian Y, Jiang L, Liang R H. Visual analytics of urban road transportation data: a survey [J]. Journal of Image and Graphics,