

# Design of Vehicle Travel Information Analysis System Based on License Plate Information

—Case Study in Guilin

Liangyu Liu, Chen Gan, Tao Wang\*, Wenyong Li

School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin Guangxi

Email: 1123860884@qq.com, \*wangtao@seu.edu.cn

Received: Oct. 18<sup>th</sup>, 2019; accepted: Oct. 31<sup>st</sup>, 2019; published: Nov. 7<sup>th</sup>, 2019

---

## Abstract

License plate recognition technology provides a convenient and efficient data source for data mining and application of traffic survey. This system is based on the license plate information acquired by the measuring points in the city. According to the demand of traffic characteristic analysis, this system mainly uses Python and SQL to carry out relevant programs' designing and data visualization processing. It realizes the generation of vehicle path, the acquisition of vehicle OD distribution between measuring points, and the generation of vehicle travel time distribution curve between neighboring points and some other practical functions. By connecting with the database, the primitive data can be updated in time and processed in large quantities, which provides a quality data source for further analysis and research of vehicle travel characteristics, provides intuitive reference for traffic planning and management, and helps the designer and management departments to carry out analysis of traffic system and design of problem-solving measures more effectively, then can improve the operation of the transportation system.

## Keywords

License Plate Information, Path Generation, OD Distribution, Visualization

---

# 基于车牌信息的车辆出行信息分析系统设计

——以桂林市为例

刘良煜, 甘 陈, 王 涛\*, 李文勇

桂林电子科技大学建筑与交通工程学院, 广西 桂林

\*通讯作者。

Email: 1123860884@qq.com, wangtao@seu.edu.cn

收稿日期: 2019年10月18日; 录用日期: 2019年10月31日; 发布日期: 2019年11月7日

## 摘要

车牌识别技术为交通调查的数据挖掘和应用提供了便捷高效的数据来源。本文基于城市各卡口测点获取的车牌信息, 结合交通特性分析需求, 主要利用Python、SQL进行相关程序设计和数据可视化处理, 实现了车辆行驶轨迹的生成、测点间车辆OD分布的获取、测点间车辆行程时间分布曲线图的生成等较为实用的功能。通过连接数据库的方式实现了对数据的及时更新和较大批量处理, 为各项车辆出行特性的进一步分析研究提供了数据源, 为交通规划和管理提供直观参考, 进而帮助设计和管理部门进行交通系统的分析和问题解决措施设计, 使交通系统的运行得到改善。

## 关键词

车牌信息, 路径生成, OD分布, 可视化

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 设计背景与目的

交通调查的主要内容包括出行起讫点调查(OD 调查)、交通量与交通设施调查、道路通行能力调查等, 而进行交通调查需要耗费大量的人力物力, 调查过程中各种不确定因素将直接影响调查结果的可信度, 所得到的数据也不具有多样性和扩展复用性, 这些都促使了现代科学技术与交通调查方法的不断融合与进步。

在交通信息化管理技术日益发展成熟的今天, 车牌识别技术也逐步完善与成熟。城市道路网络中, 各交通视频监控测点通过车牌识别技术, 可将车辆通过测定地点的时间、地点等信息存入交通信息管理中心的数据库中, 从而进一步进行数据的挖掘与分析应用。在基于车牌数据进行车流特性分析的研究工作中, 以下几项较为具有代表性:

郭昕[1]等人基于车牌识别数据, 以上海市快速路非沪车牌识别数据为例对车辆使用特征进行了多维研究, 从时间、空间、车辆属性等多个方面, 对车牌识别技术所得到的车辆出行时间与空间信息进行了深度挖掘与分析, 采用新的分析技术来对决策者与管理层进行信息反馈, 对城市交通出行特性分析来说具有重要的参考意义。马金麟[2]等人通过测点卡口获取到的车牌信息数据, 利用 VBA 编程实现对车牌数据的分析与挖掘, 获取并计算出到各个卡口间的车辆行程时间, 并在所得行程时间的基础上, 利用递归算法完成对卡口的聚类划分, 以进行交通小区的自动划分, 加以人工经验修改, 最终得到交通小区间的车辆 OD 分布矩阵。胡旭峰[3]针对现在较多应用的交通信息采集技术以及车牌识别技术所获得的信息数据特点, 分析并给出了可能出现的数据错误或缺失处理方法, 提出且验证了运用车牌识别数据来进行城市道路交通运行状态判断分析的方法, 并对行程时间的分布特性进行了研究和指标可靠性分析。

本系统基于数据库中的大量车牌信息数据, 通过对其进行清洗、分类、集成、解析, 得到城市机动

车的出行路径、OD 分布、测点间的通行时间分布等特性，并以图表、分布图等形式进行可视化呈现，提高交通调查数据的获取与车辆出行特性的分析效率，为交通的管理与设计提供直观、有效且合理的参考。

## 2. 设计方法与技术

全桂林市各测点每天收集并传回桂林市智能交通管控中心的车辆信息数据多达 150 余万条，如此庞大的数据量，只有结合数据库进行查询与分析，才能保证数据分析与处理的效率，保证结果与期望的符合度，也是更为便捷和高效的方法。

与传统的数据分析系统不同，本系统采用 Python 和 SQL 编写程序对数据库进行连接，在指定数据表中对数据进行整合分析，在保护数据的同时最大限度的保证了数据的时效性，减小了数据源与数据分析的时间差和滞后性，且运行结果将以新数据表的形式另存在数据库中，为进行各方面、多层次的后续数据挖掘提供了保障。首先对原始数据进行清洗与整合，剔除掉未能成功识别或数据严重不完整的车辆信息，以使后续分析的可靠性和准确性得到保证。其后，通过对数据表的结构进行优化，建立外键缩小数据表体积，提高数据表之间的相关性和程序的运行效率。在上述工作进行完成之后，系统便可以对数据库中的数据进行操作，实现车辆路径的查询、测点 OD 分布、OD 矩阵获取，以及相邻测点间车辆通行时间分布曲线图的生成，系统设计框架构成如图 1 所示。

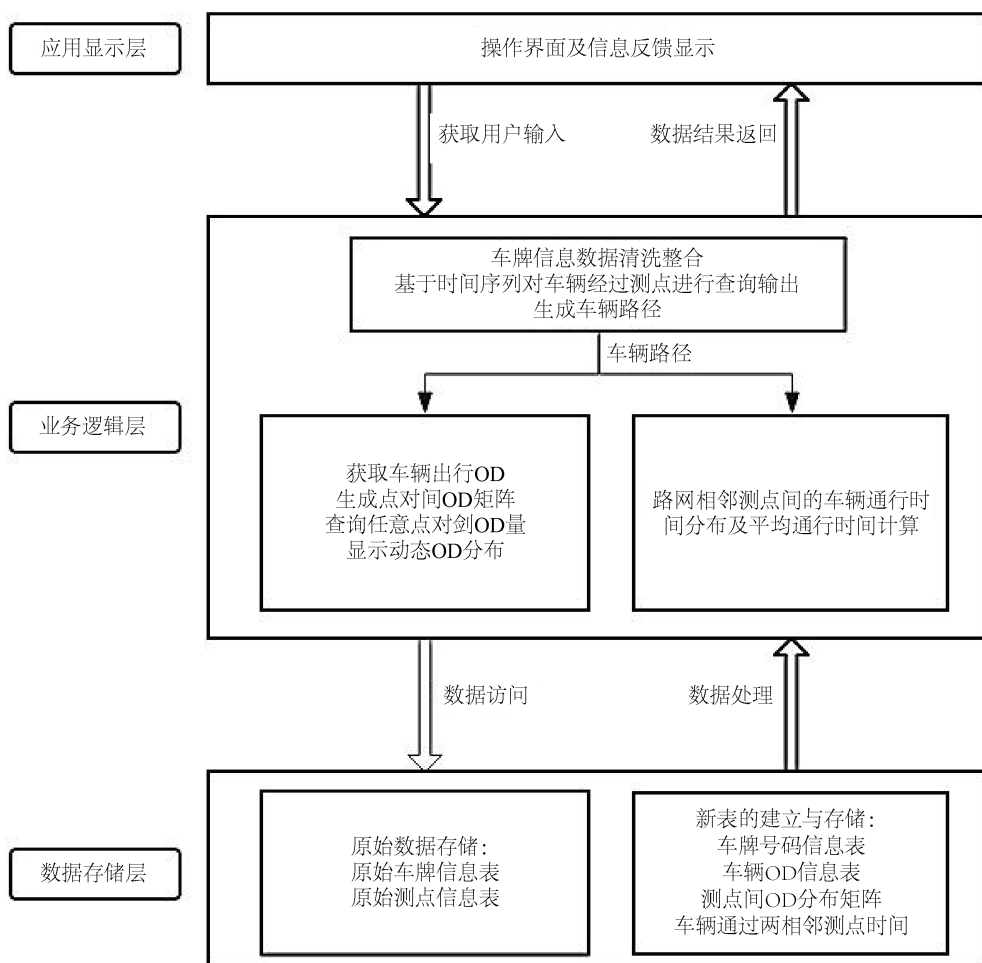


Figure 1. Framework of system design  
图 1. 系统设计框架构成

车牌信息的清洗与整合依照模糊匹配中的编辑距离算法, 利用 Python、SQL 和正则表达式完成。数据表结构的优化通过 SQL 进行多表更新和表内连接实现。车辆路径的可视化通过调用高德地图 API 接口完成。测点间 OD 分布的动态呈现以及相邻测点建车辆通行时间分布曲线的绘制则通过 Python 的 numpy 扩展库和 pyecharts 类库完成。

### 3. 数据清洗与整合

由于受到光照、车牌磨损、表面污渍等因素的影响, 车牌识别技术的读取结果中往往存在部分无效数据, 而这些没有能够正确识取的车牌信息散落在存有大量车辆信息的数据库中, 将会对系统的运行效率以及数据分析结果的可信度造成不可忽视的影响。若要保证系统运行结果的有效性, 采用车牌信息数据精准匹配的方法无疑是最佳选择, 但其直接后果是系统会对大量未能完全成功识别的车牌信息进行舍弃, 这对原始数据的完整性造成了破坏。因此, 需要依靠一种方法, 适当“放宽”对车牌数据信息完整度的要求, 在允许的阈值范围内实现对车牌数据的模糊匹配。

本系统采用编辑距离算法对数据库中存放的车牌数据信息进行模糊匹配, 以达到对现有数据较大程度的挖掘和应用。编辑距离是指将两个字符串编辑至相同所需要的最少操作次数, 从原始字符串中添加、更改、删除一个字符, 即称为一个编辑距离。故两字符串之间的编辑距离越小, 此两字符串的相似程度就越高[4]。

将相邻测点间的车牌数据两两进行匹配, 初始编辑距离设为  $d = 0$ 。由于车牌数据的字符串长度在 7 至 9 个字符之间, 且格式相对较为固定, 故可将车牌数据字符串分解为长度为 9 的元组, 依次对元组对应数位进行比较, 若字符相同则编辑距离  $d = 0$  不变, 并继续比较下一位; 若字符不同则  $d = d + 1$ , 继续进行后一位比较。计算公式如下:

$$D(A_i, B_j) = \begin{cases} 0 & i = 0, j = 0 \\ \min \begin{cases} d_{k-1} a_k = b_k \\ d_{k-1} + 1 a_k \neq b_k \end{cases} & \end{cases}$$

$$R = 1 - \frac{D(A_i, B_j)}{9}$$

式中:  $A_i$ 、 $B_j$  分别表示测点 A 的第  $i$  个车牌信息字符串元组, 测点 B 的第  $j$  个车牌信息字符串元组;  $a_k$ 、 $b_k$  分别表示元组  $A_i$  和  $B_j$  中的第  $k$  个字符;  $d_{k-1}$  表示两元组中第  $(k-1)$  个字符前的编辑距离;  $D(A_i, B_j)$  表示  $A_i$  与  $B_j$  之间的编辑距离;  $R$  表示  $A_i$  与  $B_j$  之间的相似度,  $R \leq 1$ 。

为减轻系统运算负荷, 并保证清洗后数据的可信度, 设置相似度阈值  $R_s = 0.7$ 。当  $R > R_s$  时, 可判定两车牌信息相似, 赋予相同车辆 ID; 当  $R < R_s$  时, 立即停止继续比较, 判定两车牌信息数据不可合并, 继续比较下一组车牌信息。

此次设计依据的车牌信息数据来源于桂林市智能交通管控中心的信息数据库, 所得数据覆盖桂林市 108 处卡口测点, 分布在桂林市各级城市道路、公路视频监控处。首先对原始数据中的测点名称和车牌号码进行编号, 将路网拓扑结构与车辆的通行信息数据结合起来, 以达到进行时间与空间数据挖掘的目的。采用 Python、SQL 结合正则表达式, 利用模糊匹配中的编辑距离算法, 对所得车牌数据进行清洗, 剔除掉无效信息, 并将清洗后的数据存入数据库中, 完成数据集成。数据表中具体的字段名及其含义如表 1 所示。

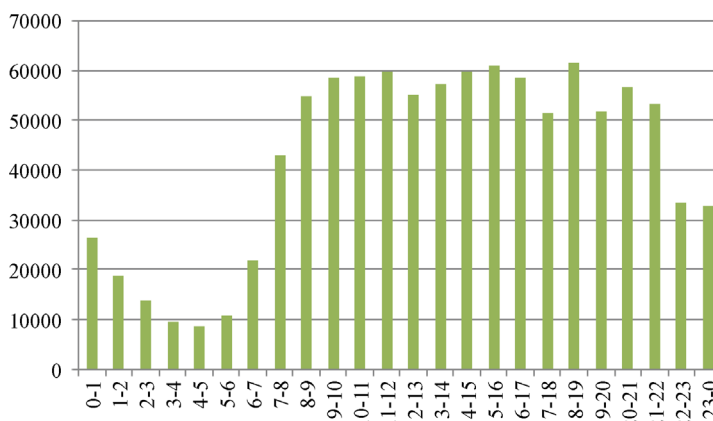
**Table 1.** Data set meaning and format**表 1.** 数据字段含义及格式

数据库字段名	含义	示例
rec_id	此条记录的编号	32
car_id	车牌号码	桂 C77777
spot_id	测点卡口编号	12
direction	车辆通行方向	北→南
pass_time	车辆经过时间	2016-10-17 08:00:00

## 4. 系统功能及应用

### 4.1. 数据描述

本文采用自广西桂林市智能交通管控中心获取的桂林市各卡口测点在 2016 年 10 月 19 日 00:00:00 至 23:59:59 所收集的全部车牌信息数据为数据源, 共计 1,017,972 条。将原始数据进行清洗整合与初步统计后, 得到桂林市在 2016 年 10 月 19 日当天的车辆出行时间分布规律, 如图 2 所示。

**Figure 2.** Guilin vehicle travel time distribution composite**图 2.** 桂林市车辆出行时间分布

### 4.2. 车辆路径查询与生成

在路径查询界面输入目标车辆的车牌号, 程序即运用 SQL 查询语句在表中进行检索, 得到该车辆在区域中的所有出行信息数据。对所获得的车辆出行数据, 以时间为序列进行排序并列表输出, 即可得到车辆的出行路径。同时调用高德地图 API 的轨迹回放和轨迹纠偏接口, 将查询获得的车辆行径测点在 HTML 文件中以路径形式呈现, 便可以完成车辆的出行路径生成与展示。示例查询结果展示截图见图 3。

### 4.3. 相邻测点间通行时间分布曲线图绘制

获取两测点间车辆的通行时间分布, 即要统计行经测点信息中包含有此两点信息的车辆个数, 并统计出通过此两测点的所有车辆所用的时间差。对数据表进行自身连接的 SQL 查询, 在 Python 语言中运用二重循环嵌套, 实现对数据的遍历, 将有效结果保存至新的数据表中, 最终得到所查询两点对间的车辆通行时间分布矩阵, 运用 numpy 扩展库进行矩阵运算处理, 借助 pandas 类库实现车辆通行时间分布图的绘制, 并计算出该两测点间所有车辆平均通行时间。具体流程如图 4 所示, 示例数据运行结果见图 5。

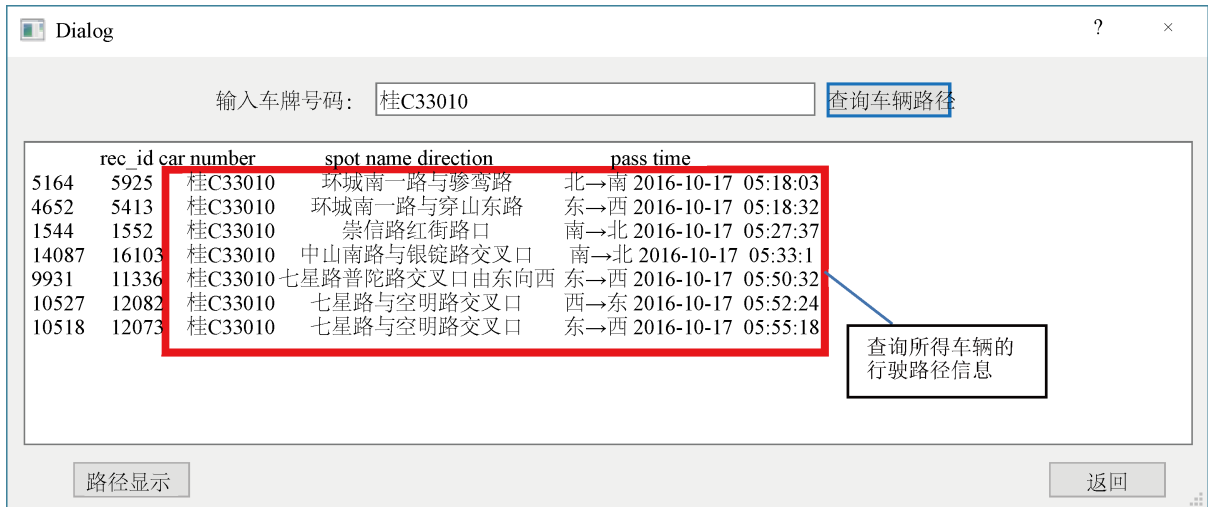


Figure 3. Example of vehicle route query results  
图 3. 车辆路径查询结果示例

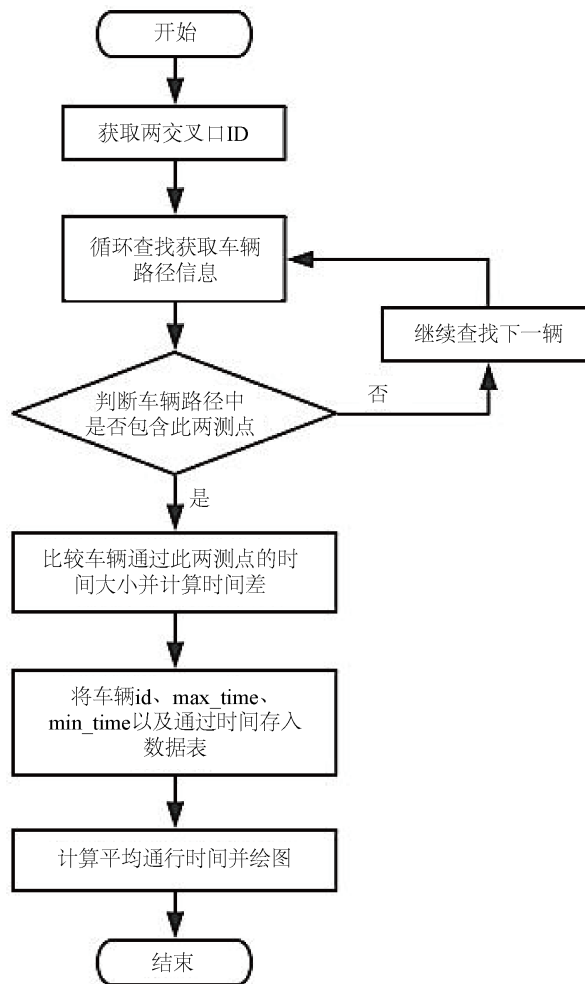


Figure 4. Flowchart of calculating travel time and drawing distribution curve  
图 4. 通行时间计算及分布曲线绘制流程图



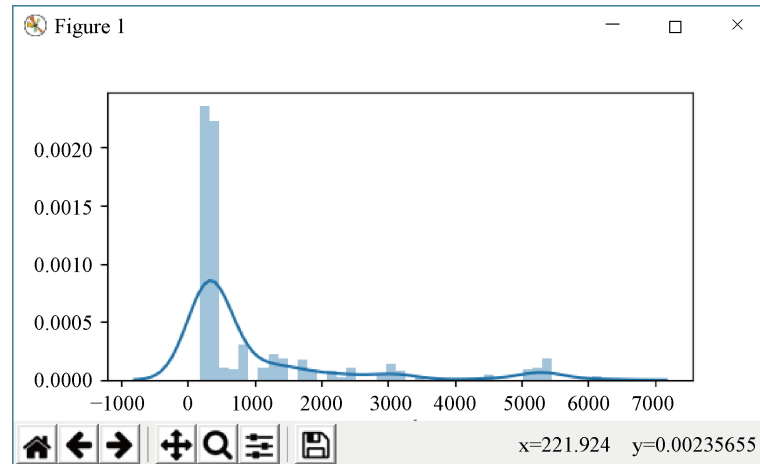


Figure 5. Curve of vehicle travel time between sample measuring points  
图 5. 示例测点间车辆通行时间分布曲线

#### 4.4. 测点间 OD 分布与矩阵生成

车辆的出行 OD 数据在交通特性分析和交通规划等方面都是必不可少的关键数据，有着极为重要的研究意义和价值。为了获取区域中各车辆的出行 OD，最基础也最为直接的思路即为先获取到所有车辆的出行路径信息，进而判断该车辆路径信息的起点与讫点。这种解决问题的思维方式映射到计算机编程中，便可表述为：获得车辆 ID 对应的、以时间为序列的测点数据，并提取出位于此项数据列表首尾的测点信息，将其保存于新建的车辆 OD 数据表中，即可完成车辆 OD 的获取。

由于要对数据表中所有车辆信息进行检索，出于对系统运行效率保证的考虑，在进行车辆 OD 查询前需对数据表进行一定的处理。通过在数据库中设置外键，得到一张测点名称及位置与 ID 的对应关系表，为每一个测点设立了专属于自身的 ID 号码，通过多表更新与表内连接，将测点 ID 存放与数据表中，对车辆路径信息进行检索排序，将所得列表数据首尾的测点 ID 以“O\_id”、“D\_id”为字段名，以车辆 ID 为主键进行排序，得到一张存放车辆 OD 信息的数据表。之后，对表中起点、讫点数据相同的车辆 ID 个数进行统计，并生成储存有测点之间车辆 OD 矩阵数据信息的 Excel 文件。与此同时，程序调用 pyecharts 将各测点间的 OD 量从文件中读取，并在 HTML 文件中进行动态地图展示，直观的将 OD 分布在区域地图中进行呈现。此部分运算过程流程图如图 6 所示。

#### 4.5. 实例应用与分析

为保证系统运行效率和数据分析结果的可靠性，结合桂林市车辆出行时间分布特征，选取桂林市 2016 年 10 月 19 日早 7 时至 8 时之间各测点采集到的车牌信息进行车辆 OD 分布的分析和具体应用，此时段即将迎来出行早高峰，具备高峰期交通流分布的初步趋势又没有大量的交通拥堵干预分析结果，是较为合理的数据分析对象。分析最终得到的车辆 OD 分布结果如图 7、图 8 所示。

由宏观结果可知桂林市车辆 OD 分布主要以秀峰、叠彩、象山、七星、燕山五大辖区为中心向周边各县区辐射。微观来看，主要的 OD 点分布在环城北一路、环城北二路、普陀路、漓江路、环城西二路所围成的桂林市老城中心区内部，且具有较高的聚合性。选取桂林市老城中心区为研究区域，结合车辆 OD 分布结果，对老城中心区进行交通小区的划分，进一步对现状 OD 分布进行分析。

结合桂林市路网规划及主要的土地利用性质分布(图 9)，对桂林市老城中心区进行交通小区的划分，并通过对各小区内测点信息的整合得到各交通小区的车辆 OD 分布矩阵。小区划分示意图见图 10，各小

区的交通发生吸引量统计见表 2。

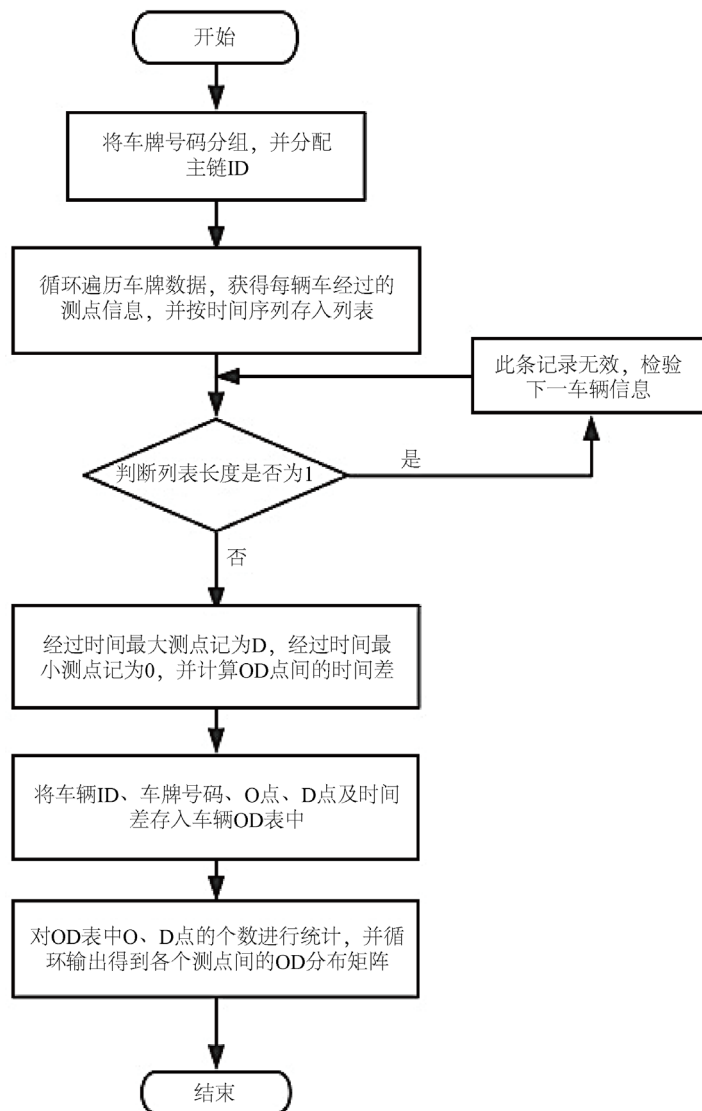


Figure 6. Flowchart of acquisition of OD matrix

图 6. OD 矩阵获取流程图



Figure 7. Macroscopic and mesoscopic screenshots of dynamic OD distribution among measuring points

图 7. 测点间 OD 分布动态示例结果宏观及中观截图



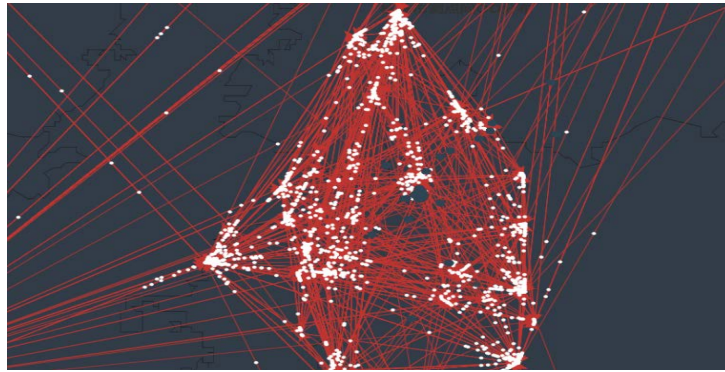


Figure 8. Microscopic screenshots of dynamic OD distribution among measuring points

图 8. 测点间 OD 分布动态示例结果微观截图

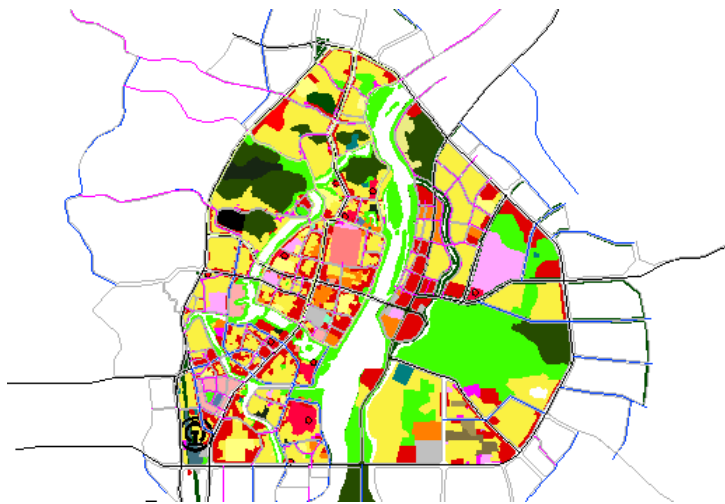


Figure 9. Distribution of land use property and road network in the central area of Guilin

图 9. 桂林市老城中心区土地利用性质及路网分布



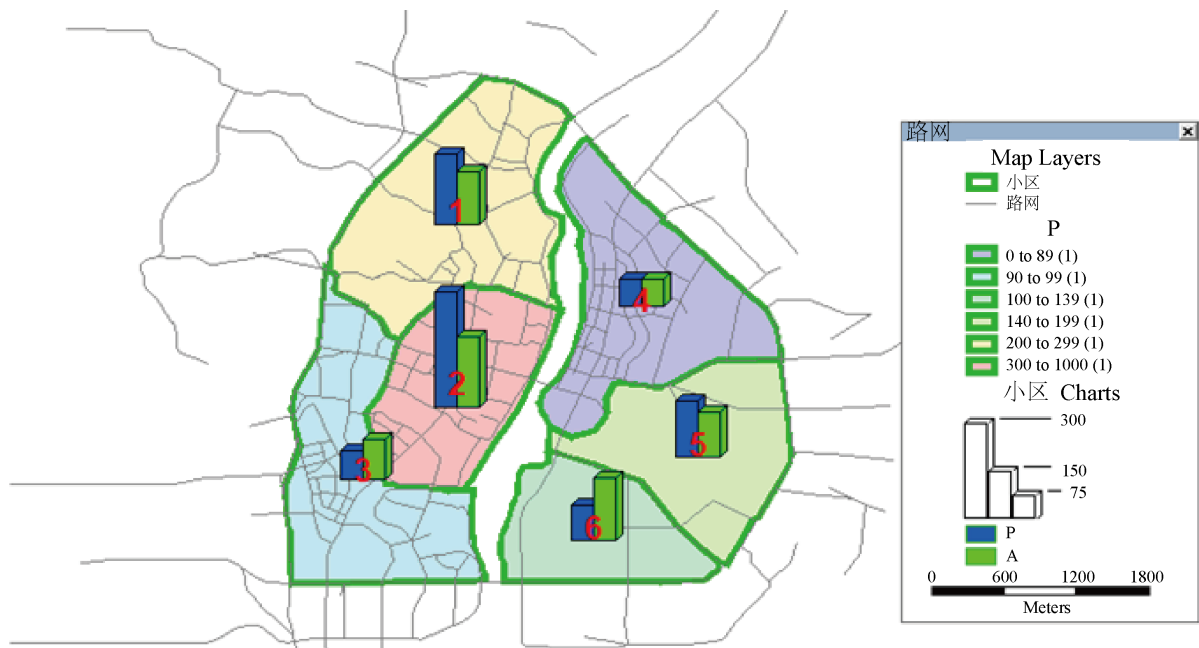
Figure 10. Distribution of traffic district in the central area of Guilin

图 10. 桂林市老城中心区交通小区划分

**Table 2.** Measurement points and occurrence attraction in traffic districts (times/hour)  
**表 2.** 小区所含测点及发生吸引量(次/小时)

交通小区	所含测点编号	小区发生量	小区吸引量
1	28, 31, 43, 47, 96, 97, 98	219	167
2	23, 25, 26, 27, 39, 48, 61, 101, 102, 103	361	222
3	22, 35, 41, 65, 66, 86, 88	94	126
4	24, 40, 76, 77, 94	87	87
5	53, 54, 56	178	145
6	71	113	194

依据现有数据, 采用双约束重力模型进行交通出行分布预测, 在 TransCAD 中将各小区质心间的距离最短路作为各小区间的阻抗矩阵, 平衡后的各交通小区 PA 量分布如图 11 所示。



**Figure 11.** Distribution of balanced traffic production and attraction in each district  
**图 11.** 平衡后各小区交通发生吸引量分布

结合外部小区以及现状各交通小区的 PA 分布, 用重力模型对将来年桂林市老城中心区 OD 分布进行预测, 并绘制期望线如图 12 所示。

通过对现状 OD 分布以及预测 OD 期望线的比较分析, 可以看出 2 号交通小区内车辆出行发生和吸引量明显高于其他几个小区, 4 号小区的 OD 点分布也较为密集, 这与两个交通小区的土地利用性质和特殊地理位置关系密切。商业服务区和市民居住地在该两个小区密集分布, 其中包括了中心广场、象山公园、工人文化宫、东西巷、七星公园等重要的商旅服务业分布, 是重要的交通发生和吸引点。且自中山中路开始, 经过十字街、解放桥、栖霞路、六合路, 共同构成了连接漓江的东西两岸的重要交通干道, 承载过江的绝大部分交通流。

基于此前经数据分析得到的 OD 分布以及 PA 量, 对桂林市老城中心区进行交通分配(结果如图 13 所示), 可发现自解放桥经自由路、东江二路、栖霞路到六合路路段为拥堵路段。其中, 解放桥为连接东西

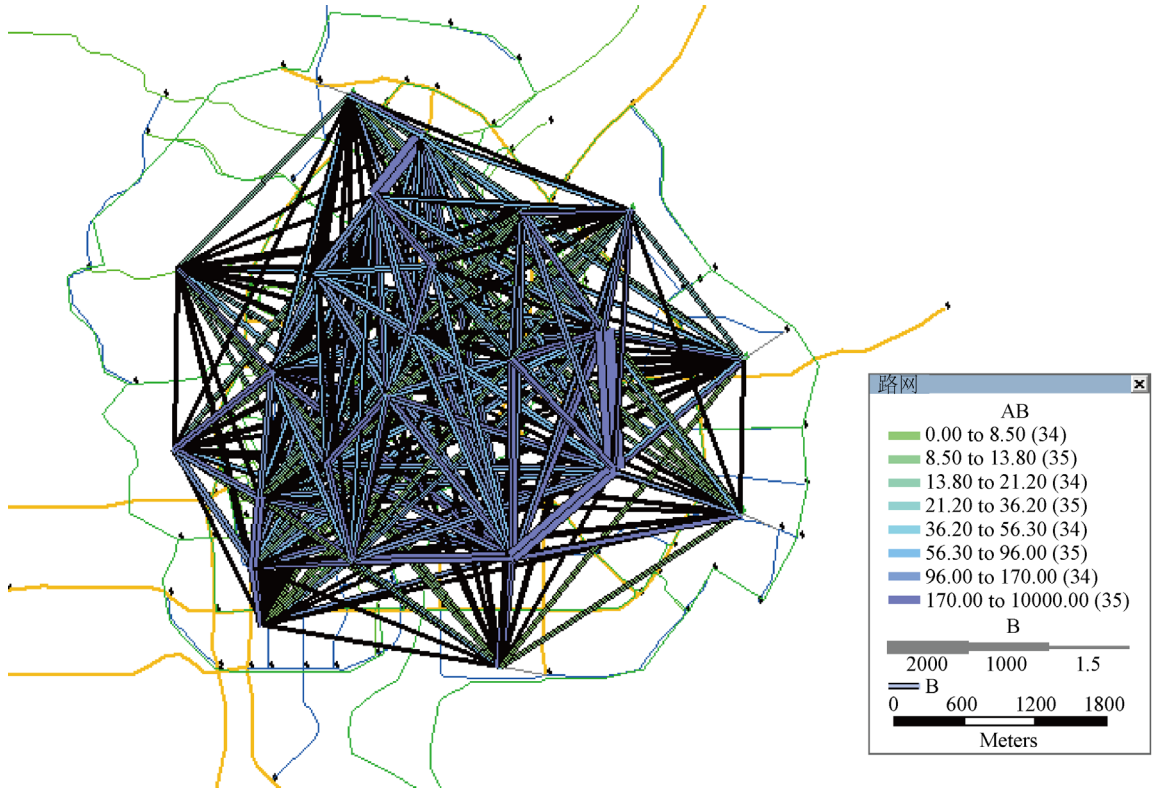


Figure 12. OD expectation line for vehicle travel in the central area of Guilin

图 12. 桂林市老城中心区车辆出行 OD 期望线

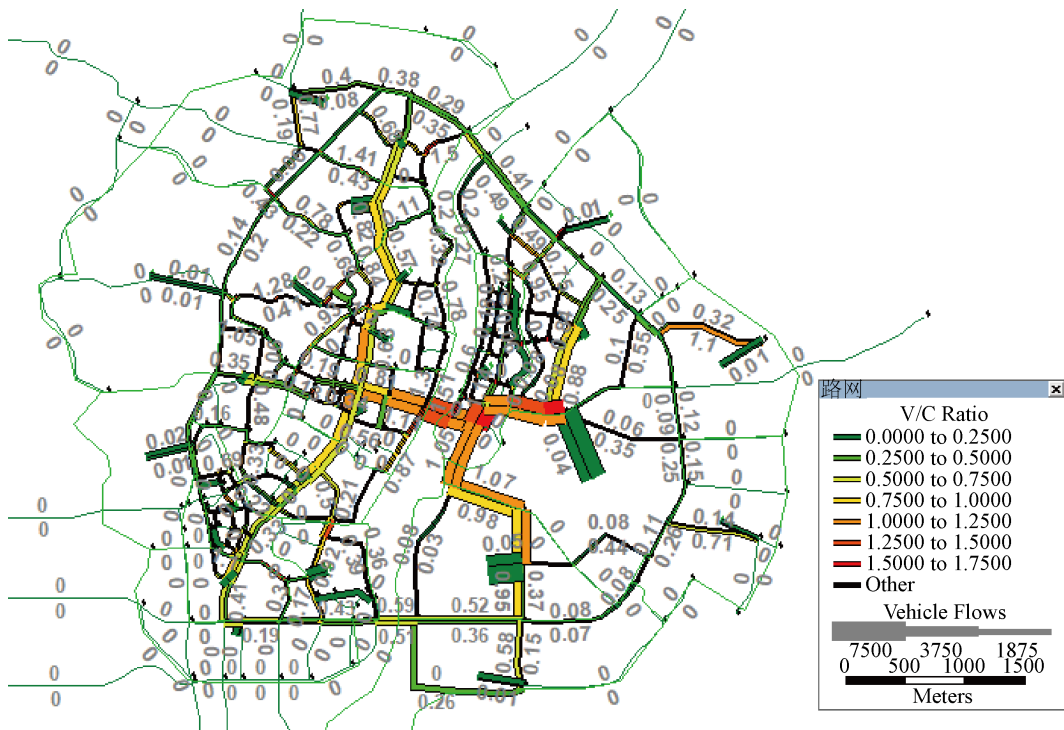


Figure 13. Results of traffic distribution section in the central district of Guilin

图 13. 桂林市老城中心区交通分配路段饱和度结果

区域的主干道, 交通压力大, 经常发生拥堵。六合路为连接城区中心与环城路的一条重要道路, 进入和驶离市中心区的车辆有大量会经过六合路至栖霞路一段, 而由于小东江上的栖霞桥路幅所限, 导致六合路、栖霞路的通行能力严重不足, 拥堵也成为常态。

## 5. 功能扩展

为体现系统的功能性与实用性, 本系统内置了小型 OD 矩阵预测计算功能。将现有的交通小区 OD 矩阵数据文件以 CSV 格式导入, 并设置目标年交通发生与吸引力, 系统将运用预先设置好的福莱特(Fratar)法对各交通小区将来年的 OD 分布进行预测及收敛判别, 并以 CSV 文件另存为系统根目录中, 便于后续操作(程序预设的收敛标准为  $\varepsilon = 3\%$ )。

本系统的设计初衷是便于交通调查数据的获取与车辆出行特性的分析。通过对数据库中存放的数据进行清洗、集成、解析, 以获得车辆短时出行特性, 包括车辆行驶路径、测点间 OD 分布、相邻测点间车辆通行时间分布等。若在条件允许的情况下, 与智能交通管控中心的数据库进行连接, 即可为系统提供大量实时的数据来源, 系统的运行结果将会进行实时更新, 这为进行道路交通运行状态判别、道路通行状态短时预测等研究提供了具有参考价值的基础数据[5]。

## 6. 总结与展望

通过对桂林市智慧交通管理中心的车牌识别信息进行数据挖掘, 从中对桂林市车辆交通运行状态进行部分特性的分析, 并进行了分析数据的可视化呈现, 减少了人员的调动和前期准备, 使交通调查在人事物力上的消耗都有了很大程度的降低, 提高了交通数据获取与分析整个过程的效率。程序支持数据的可视化呈现, 可以方便决策者直接形象的了解交通系统中车辆的出行特征, 为决策者更好地进行道路规划设计、交通规划与管理提供了有效的帮助。系统所需数据可从中心库中拷贝, 或直接连入数据库进行创建查询等操作获取, 运用程序进行分析, 十分高效便捷, 保证了研究人员可以将更多的精力投入数据的挖掘与信息获取中。同时, 实现了路段车辆通行时间分布曲线的生成, 对进一步判别道路运行状态和短时预测提供了数据来源与基础。

## 基金项目

广西高校大学生创新创业项目(No. 201710595051)。

## 参考文献

- [1] 郭昕, 陈川. 基于车牌识别数据的车辆使用特征研究: 以上海市快速路非沪车牌识别数据为例[J]. 综合运输, 2016(1): 77-84.
- [2] 马金麟, 张宗博, 谢君平, 等. 基于车牌识别数据的车辆 OD 矩阵获取研究[J]. 重庆理工大学学报(自然科学), 2017, 31(7): 48-55.
- [3] 胡旭峰. 基于车牌识别数据的城市道路交通状态判别及旅行时间可靠性的研究[D]: [硕士学位论文]. 青岛: 青岛科技大学, 2017.
- [4] 刘静. 基于模糊匹配的多级车牌匹配技术[J]. 现代电子技术, 2015(7): 135-138.
- [5] 王亚群. 基于车牌识别数据的城市路网运行状态识别方法实践[J]. 交通与运输(学术版), 2018(z1): 94-98.