|  |  |
| --- | --- |
| 学号 | 2013301500148 |
| 密级 |  |

**武汉大学本科毕业论文**

|  |
| --- |
| **城市租赁车数据可视化数据**  **分析系统研究与实现** |

|  |  |
| --- | --- |
| 院（系）名 称 ： | 计算机学院 |
| 专 业 名 称 ： | 计算机科学与技术 |
| 学 生 姓 名 ： | 黄上佛 |
| 指 导 教 师 ： | 杜博文 讲师 |

二〇一七年五月

**郑 重 声 明**

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名： 日期：

摘 要

关键词：关键词1；关键词2；关键词3；……（3-5个词）

**ABSTRACT**

**Key Words**: Keyword1; Keyword2; Keyword3; ……

目 录

[1 □□□□ 1](#_Toc468699222)

[1.1 □□□□□□ **错误!未定义书签。**](#_Toc468699223)

[1.2 □□□□□ **错误!未定义书签。**](#_Toc468699224)

[1.3 □□□□□ **错误!未定义书签。**](#_Toc468699225)

[1.3.1 □□□□□□□ **错误!未定义书签。**](#_Toc468699226)

[1.3.2 □□ **错误!未定义书签。**](#_Toc468699227)

[1.4 □□□□ **错误!未定义书签。**](#_Toc468699228)

[2 □□□□□□□ 1](#_Toc468699229)

[2.1 □□□□□□ **错误!未定义书签。**](#_Toc468699230)

[2.2 □□□□□□ 6](#_Toc468699231)

[参考文献 7](#_Toc468699232)

[致谢 12](#_Toc468699233)

[附录 13](#_Toc468699234)

# 绪论

## 背景与意义

随着经济形势和国家产业结构的变化，人们对拥有汽车的观念也在变化。越来越多的家庭、个人开始尝试租用各种汽车。人们对汽车的功能和配置，在某一特定时间内也有了特殊的需求，越来越多的中小企业也利用租赁汽车来满足其对汽车的需要。汽车租赁业在这样的背景下迅速发展起来。近年来，我国汽车租赁业发展也非常迅速，并逐步形成规模。但租赁车行业在我国是一种新型的服务行业，政府对该行业的管理和领导还处于探索阶段。

租赁车规模不断增长的同时，其产生的数据也不断增多，高效的运用这些海量数据，将对汽车租赁行业的发展决策有指导性的意义。第一，可以通过各个公司的备案数据等掌握整个租赁行业的发展现状以及发展趋势；第二，汽车租赁分为多种租赁模式（短租，长租，分时租赁等），通过分析各种租赁模式下车辆的运行特征，可以确定租赁车在城市公共交通所起的作用；第三，通过分析各租赁公司上传的数据来评价其服务水平，政府部门可以对不同服务水平的公司采取相应的奖惩措施；第四，通过分析租赁车的交通出行量（Origin-Destination），可以反映租赁车在空间上的分布情况，为租赁车的站点选址和规模预测提供可靠的数据支持。

然而原始租赁车数据一般由杂乱无章的数据构成，如何快速、直观的分析这些数据就是一个亟待解决的问题。基于这个问题，本文建立的是一个城市租赁车数据的可视化分析系统，采用可交互的形式对租赁车数据进行动态分析。

## 相关说明

本文是对城市租赁车数据进行分析研究，在此对数据来源以及租赁车相关的概念进行说明。

### 数据来源

本文来自于实际课题，所使用的租赁车数据来自于北京市租赁车行业实际生产环境，保证数据的真实性。目前获取到的租赁车数据为2016年各月份的数据，由于数据各月上传的波动性大，为了保证系统分析的可靠性，本系统使用数据质量较好的8-11月份的数据。

### 租赁模式的定义

根据原始的租赁合同信息，将租赁车辆分为分时租赁、普通租赁短租、普通租赁长租。具体区分方法见表1.1

|  |  |  |
| --- | --- | --- |
| 表1.1 租赁模式区分方法 | | |
| **序号** | **租赁模式** | **区分方式** |
| 1 | 长租 | 车辆租赁时长大于30天 |
| 2 | 短租 | 车辆所属租赁公司无分时租赁业务，且租赁时长小于30天；车辆所属租赁公司有分时租赁业务，且租赁时长大于1天、小于30天 |
| 3 | 分时租赁 | 车辆所属租赁公司有分时租赁业务，且合同租赁时长小于24小时 |

其中，分时租赁是短租的一种特殊模式

### 租赁车相关数据

租赁合同：租赁公司与租赁人签订的合同数据，包括车牌号，租赁公司，租赁人，租赁时长等。

备案数据：租赁车辆在北京市交通委汽车租赁处的备案数据，包括公司租赁ID，公司名，地址，车牌号，备案人，备案时间等。

车辆GPS数据：北京市汽车租赁行业管理与服务信息系统接收到的租赁车辆的日行驶轨迹GPS信息，包括车牌号，GPS生成时间，经纬度。

为了分析方便，以上数据均已转成CSV（逗号分隔值）格式。

### 租赁车OD

OD（Origin-Destination）指的是起终点的交通出行量，租赁车OD表示租赁车起止点的分布情况，即出发地数据量和目的地数据量。

### 交通小区

为了减少交通系统控制与管理的复杂性，将整个城市道路网络分为不同的交通子区，这些交通子区称为交通小区。根据不同的需求可以有不同的划分方法，本系统使用的是北京交通委划分的461交通小区，将北京换分成461块交通小区。交通小区部分字段如表1.2

|  |  |  |
| --- | --- | --- |
| 表1.2 交通小区部分字段 | | |
| 字段 | 字段说明 | 例子 |
| id | 交通小区的唯一标识 | 1 |
| name | 交通小区名字 | “王府井” |
| centerX | 小区中心经度 | 116.40175273139036 |
| centerY | 小区中心纬度 | 39.9693391035951 |

## 论文组织结构

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

# 系统功能

本系统为了对城市租赁车数据进行一个比较全面的分析，分为四个功能：租赁行业的数据概况展示，租赁车辆统计特征分析，各公司的GPS数据质量分析，租赁车OD情况分析。

## 租赁行业的数据概况展示

该功能包含以下三个部分

1. 基础数据：基础数据主要为8月-11月的租赁合同数、备案数以及上传了GPS的车辆数。通过立体柱状图的形式展现三个数据在各个月的上传情况，以及增减趋势。
2. 三种租赁模式车辆占比情况：使用南丁格尔玫瑰图（特殊的饼图）生动的展示三种模式的车辆占比，并且可以选取月份来查看不同月份的占比情况。
3. GPS数据随时间变化曲线： 使用动态曲线呈现三种租赁模式在8-11月每天GPS上传的情况。

## 租赁车辆运行特征分析

该功能根据已有的车辆数据，对不同模式车辆从租赁时长、使用强度、日均行驶里程、出行时间四种指标进行运行特征分析。

1. 租赁时长：对租赁合同进行统计分析，来说明各个模式中租赁时长的分布情况，从而分析各模式的主要功能以及服务的对象。
2. 使用强度：使用强度指的是租赁用户平均每天使用租赁车的次数。该特征主要统计三种租赁模式下在工作日和双休日的使用次数情况。
3. 日均行驶里程：通过统计各种租赁类型每日的行驶里程，找到不同模式下车辆行驶里程分布的异同点，和相同模式下工作日和双休日里程分布的不同之处。
4. 出行时间：通过统计各租赁模式下工作日和双休日的出行时间分布，掌握各类型车辆是否具有明显的通勤特征。

## 各公司的GPS数据质量分析

按照规定，所有租赁公司的租赁车必须按照规定上传GPS数据。因此，GPS数据质量是评价一个租赁公司优劣的重要指标。本功能按照GPS的数据质量对各个公司进行评比，选出排名前5和倒数前5的公司。本系统对GPS的质量采用三个评价指标：

1. 重复数据上传数占比：重复率越高，GPS质量越差
2. 高频活跃车辆数占比： GPS上传时间间隔小于2分钟的车辆定义为高频活跃车辆，一天中高频活跃车辆越多，车辆行驶轨迹的可重塑性越高。
3. GPS密度：一辆车上传的相邻2个GPS距离小于1000m的定义为高密度GPS，GPS密度越大，GPS质量越好。

评选出优质和劣质公司后，用户可以选取某个公司，在地图上查看该公司在某一天某辆车的GPS轨迹。

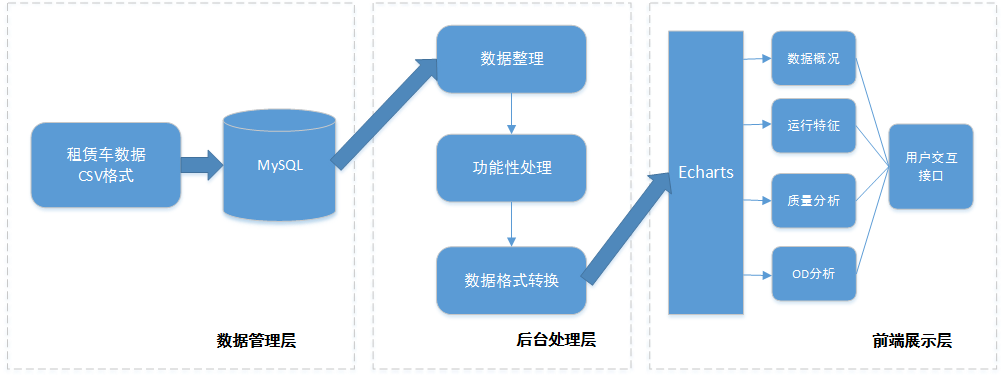
## 租赁车OD情况分析

根据租赁车GPS数据，将所有租赁车的出发地和目的地划分到相应的交通小区内，然后统计每天在不同时间段下各个交通小区的出发量、到达量和吞吐总量，再对这个三个量进行排序。用户可以选择日期，系统就可以在地图动态地展示租赁车在一天内的迁徙情况，并且显示交通量最大的前5名交通小区。政府部门和租赁公司根据这些数据，可以在交通量大的地方设置租赁站点，满足城市居民的租赁需求。

# 系统设计

## 系统框架

图4.1 系统框架



系统框架如图4.1所示，租赁车可视化数据分析系统主要分为三层结构：数据管理层、后台处理层、前端展示层。

数据管理层：租赁车数据基本上都是由CSV格式的数据组成，该格式全为逗号分隔，虽然简单，但是数据的结构性和关联性不强，当进行数据处理时不方便，因此需要将CSV格式的数据以表的形式保存在MySQL数据库中，并且在保存的过程中，还可以对一些无关紧要的数据进行剔除，使数据更加清晰。

后台处理层：本系统的后台使用JAVA开发，并采用SSM架构（详见3.2）。首先，使用SQL语句从MySQL中读取所需要的数据，并对数据进行相应的整理。然后根据不同的功能，对数据进行相应的功能性处理，如统计、排序等。最后，将处理好的数据转换成JSON(JavaScript Object Notation)格式，并形成数据接口，提供给前端访问。

前端展示层：该层即为WEB页面，由jsp+css+js构成。由于需要对租赁车数据进行可视化分析，本系统采用强大的Echarts图表库（详见3.3）进行展示。前端由4个web页面组成：数据概况，运行特征，OD分析，分别对应第2章描绘的4个功能。每个页面都有许多简单的组件提供给用户操作，实现系统的可交互性。

## 后台框架——SSM

## 可视化工具——Echarts

# 具体实现

## 数据概况

## 统计特征

## 质量分析

## OD分析

# 实验结果与分析

# 总结与展望

参考文献

[1] 戴军, 袁惠新.膜技术在含油废水处理中的应用[J].膜科学与技术，2002，22(2)：59-64

[2] 毛侠，孙云.和谐图案的自动生成研究[A].第一届中国情感计算及智能交互学术会议论文集[C].北京：中国科学院自动化研究所，2003：277-279.

[3] 王湛.膜分离技术基础[M].北京:化学工业出版社，2000：14-21，30.

[4] 张志祥. 间断动力系统的随机扰动及其在守恒律方程中的应用[D].北京:北京大学数学学院,1998.

[5] World Health Organization. Factors regulating the immune response: report of WHO Scientific Group[R]. Geneva: WHO, 1970.

[6] 河北绿洲生态环境科技有限公司.一种荒漠化地区生态植被综合培育种植方法:中国,01129210.5[P].2001-10-24.

[7] GB/T16159-1996,汉语拼音证词法基本规则[S].北京:中国标准出版社,1996.

[8] 毛侠.情感工学破解“舒服之谜”[N].光明日报，2004-04-17(B1).

[9] 陈剑.上博简《民之父母》“而得既塞於四海矣”句解释[EB/OL].简帛研究网站，http://www.bamboosilk.org/Wssf/2003/chenjian03.htm．2003-01-18

致谢

感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢

感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢感谢

附录

□□□□□□□根据实际情况更新,若分附录A、附录B……，就把上面标题添个“A”，再来附录B…… 无附录就删去本页。附录其它标题格式注意遵循《规范》