|  |  |
| --- | --- |
| 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 □□□□□□ **错误!未定义书签。**](#_Toc468699231)

[参考文献 5](#_Toc468699232)

[致谢 29](#_Toc468699233)

[附录 30](#_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 | 样例 |
| 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、出行时间四项统计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。

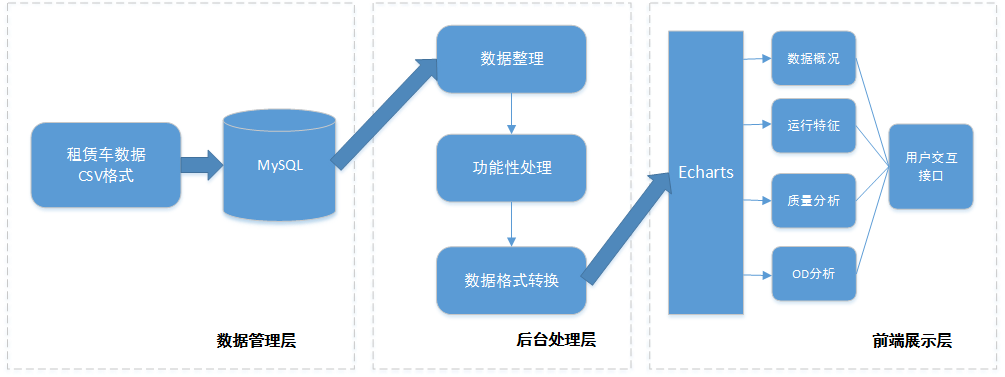
## 本章小结

本章主要介绍了系统的几大主要功能。2.1节介绍了租赁行业的数据概况展示，主要展示租赁行业的数据规模，发展现状以及发展趋势。2.2节根据四项指标对租赁车辆的运行特征进行分析，通过租赁模式间的横向对比以及工作日与双休日的纵向对比，可以掌握各类型租赁车辆的异同。2.3节介绍了根据三个评价指标对各公司上传的GPS数据进行质量分析，从而对公司进行评价。2.4节介绍了租赁车OD情况分析，通过交通小区的交通量，可以为租赁车的站点设置提供数据支持。

# 系统设计

## 系统框架

图.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3.3.1）。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。前端使用AJAX引擎，调用后台提供的接口，获取相应数据。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3.4.1）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。

## 数据库设计

由于租赁车数据量大，各种数据结构不一致，为了更好地对租赁车数据进行分析，本系统采用MySQL数据库管理系统，对数据进行表格化管理。

### MySQL

MySQL是一个小型的关系型数据库管理系统，由瑞典MySQL AB公司开发，现今被Sun公司收购。它拥有许多优点，而且开放源代码，因此是很多中小型企业和个人开发者的首选。

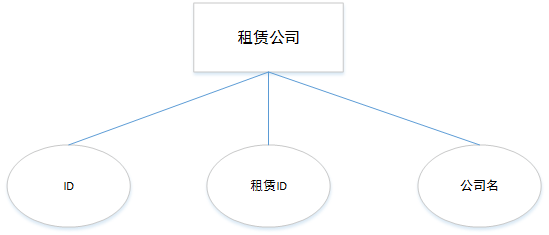
1. 支持常见的SQL语句规范
2. 具有良好的运行效率
3. 可以处理千万条记录的大型数据
4. 可移植性高，安装简单小巧
5. 调试、管理以及优化非常简单

### 实体ER图

为了进行高效的数据分析，需要将租赁车数据设计成关系型数据表，以下是主要数据的ER图（实体-联系图）。

1. 租赁公司

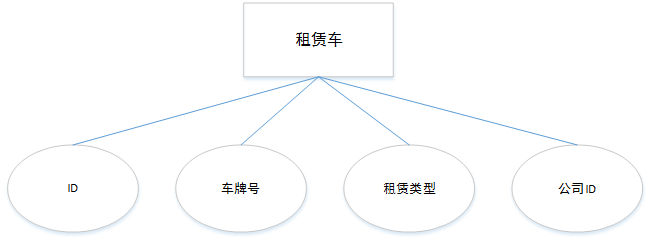
图.2 租赁公司ER图



如图3.2所示，租赁公司实体主要有三个字段构成：ID，租赁ID，公司名。其中ID为公司在数据库中的标识，租赁ID为公司在备案数据中的ID。

1. 租赁车

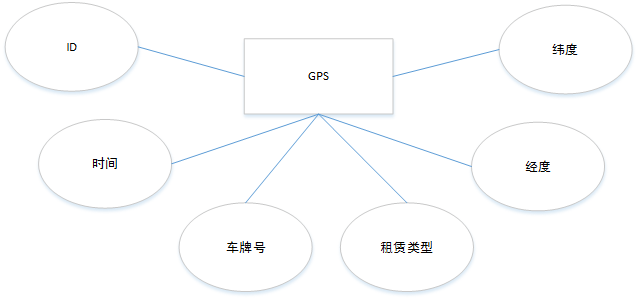
图.3 租赁车ER图



如图3.3所示，租赁车实体由四个字段构成：ID，车牌号，租赁类型，公司ID。租赁类型有长租、短租、分时租赁三种，分别取值1、2、3。根据公司ID，可以找到车辆对应的公司。

1. GPS

图.4 GPS ER图



如图3.4所示，GPS由6个字段构成：ID，时间，车牌号，租赁类型，经度，纬度。根据车牌号可以找到对应的车辆。

1. 公司评价指标

图.5 公司评价指标 ER图



如图3.5所示，公司评价指标由四个字段构成：公司ID，重复率，高频活跃车辆比，GPS密度。该实体主要功能就是为了实现第三个系统功能——各公司的GPS数据质量分析，实体的后面三个字段均为评价GPS数据质量的指标。

## 后台设计

### 后台框架SSM

SSM英文全称为Spring+SpringMVC+MyBatis，由Spring、SpringMVC、MyBatis三个开源框架整合而成的，是一种轻巧简单的web项目框架。

Spring是一个轻量级的分层框架，可以将系统进行模块化管理。

SpringMVC是Spring框架的后续产品，已经融合在Spring Web开发里面。主要采用经典的MVC架构，定义了控制器（Controller）、视图（View）、模型（Model），用业务逻辑、数据和界面显示分离的方法组织代码，使得系统的输入、处理和输出功能结构清晰，分工明确。

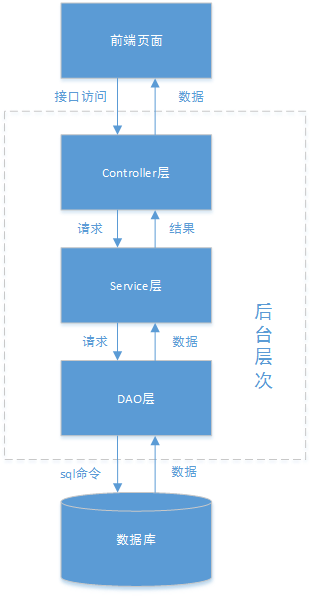
MyBatis是一个基于Java的持久层框架，它使用简单的XML语言和注解将Java的对象映射成数据库中的记录，将SQL语言嵌入在XML中，摆脱原本复杂的JDBC代码和参数的手工设置，使数据库的增删改查变得非常简单。

本系统的后台采用SSM框架，对后台功能进行分层管理。

### 后台层次结构

后台采用MVC的设计理念，将后台分成三个层次：DAO层，Service层，Controller层，结构如图3.6所示。

图.6 后台层次结构图



1. DAO层：该层主要是做数据持久层的相关工作，负责与数据库相互联络的任务都在这。DAO层首先根据需要，设计相应的数据访问接口，如查询、插入等，然后使用MyBatis规范，用XML格式的SQL语句实现各接口。当Service层调用DAO层某个接口时，会直接对数据库执行相应的SQL命令。MyBatis会根据数据库中表，自动生成与表相对应的Java实体类，因此DAO层和数据库进行交互时，不需要进行数据类型的转换。
2. Service层：该层主要负责业务模块的逻辑应用设计。首先设计业务接口，再根据业务逻辑设计其实现类。它的业务逻辑的实现，需要调用DAO层的数据接口，对数据进行加工以后，将结果提供给Controller层。Service层是整个后台的核心内容，是数据处理的关键步骤。通过封装该层的业务逻辑，有利于通用逻辑的独立性和重复利用性，使结构更加简洁，层次更加清晰。
3. Controller层：该层负责具体模块流程的控制，通过调用Service层的接口来完成模块的功能。针对具体的业务流程，需要设计不同的控制器，在设计过程中可以将流程进行抽象归纳，设计出可以重复利用的子单元流程模块。在SpringMVC中，定义一个Controller非常简单，只需使用@Controller标记一个类，然后通过@RequestMapping 和@RequestParam 等一些注解定义URL 请求和Controller 方法之间的映射，就可以被外界（前端）进行访问。

## 界面设计

本系统的功能主要通过web页面进行展示，主要包含4个页面：数据概况，运行特征，OD分析。为了提高界面的美观性和系统的可交互性，本系统使用了百度公司的图表库Echarts，用图表的形式将结果呈现出来。

### 可视化工具——Echarts

Echarts，Enterprise Charts的缩写，商业级数据图表，一个纯Javascript的图表库，可以流畅的运行在PC和移动设备上，兼容当前绝大部分浏览器（IE6/7/8/9/10/11，chrome，firefox，Safari等），底层依赖轻量级的Canvas类库ZRender，提供直观，生动，可交互，可高度个性化定制的数据可视化图表。创新的拖拽重计算、数据视图、值域漫游等特性大大增强了用户体验，赋予了用户对数据进行挖掘、整合的能力。

Echarts功能强大，提供了丰富的图表类型，如常规的柱状图、折线图、饼图、散点图，用于统计的盒形图，用于地理数据可视化的地图、线图、热力图等等。图3.7为Echarts的整体结构

图.7 Echarts整体结构图



### Web页面

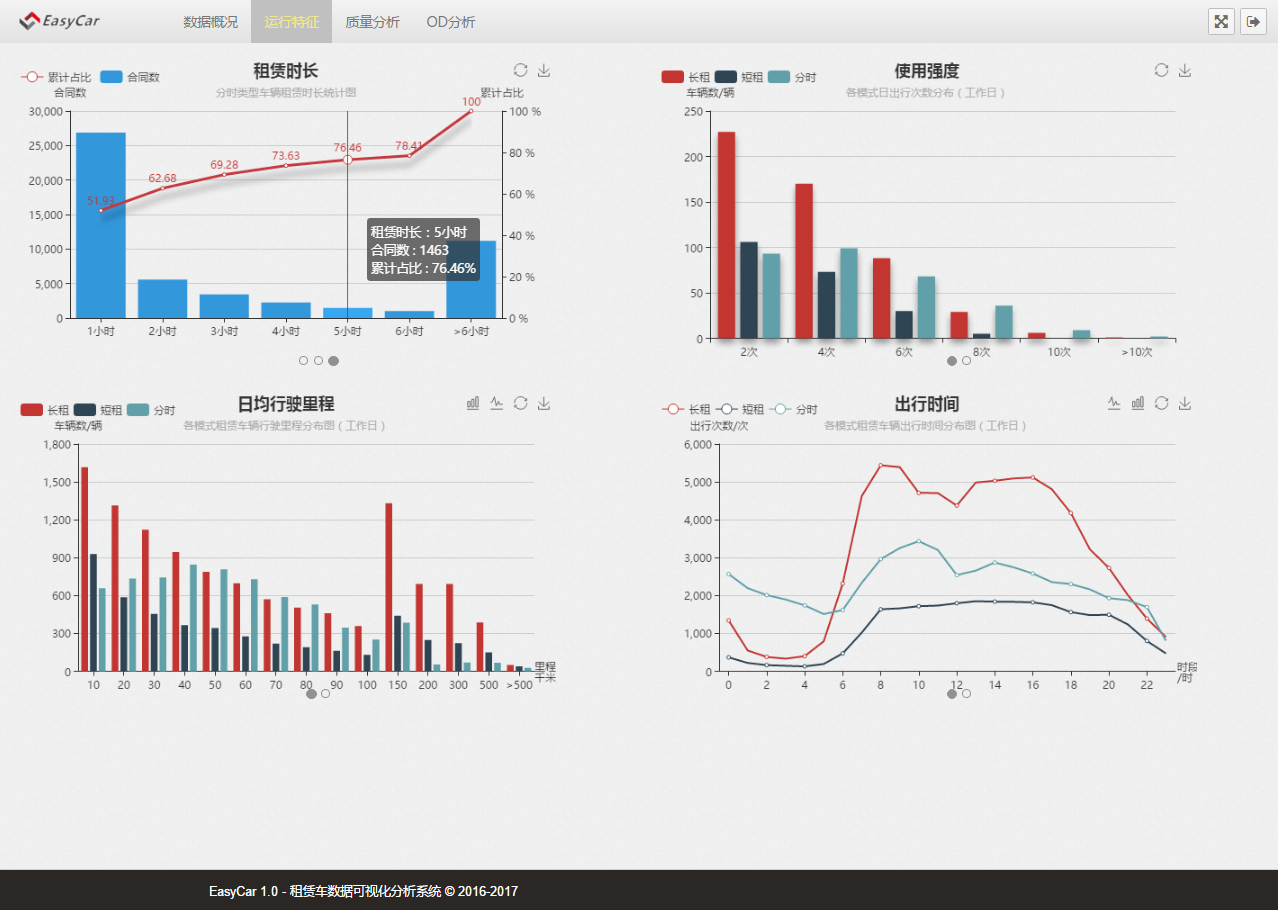
1. 数据概况页面：该页面主要展现租赁行业的数据概况，包括基础数据，各租赁类型上传的GPS车辆数，以及GPS数据量随时间变化曲线。页面如图3.8所示

图.8 数据概况页面



1. 运行特征页面：该页面主要展示租赁车辆运行特征分析功能，包括租赁车的四个统计特征，租赁时长、使用强度、日均行驶里程、出行时间。页面如图3.9所示。

图.9 运行特征页面



1. 质量分析页面：该页面主要展示GPS数据质量分析功能，根据三种评价指标，对各公司进行评价排序，并且在地图上可以查看某公司某辆车的GPS轨迹。页面如图3.10所示。
2. OD分析页面：该页面主要展示租赁车的OD情况，根据租赁车的OD交通量对交通小区进行排序，点击不同的排序方式，选择日期，就可以在地图上查看该排序方式下租赁车当天24小时的迁徙动态图，并显示出发量、到达量和吞吐量排名TOP5的交通小区。页面如图3.11所示。

图.10 质量分析页面

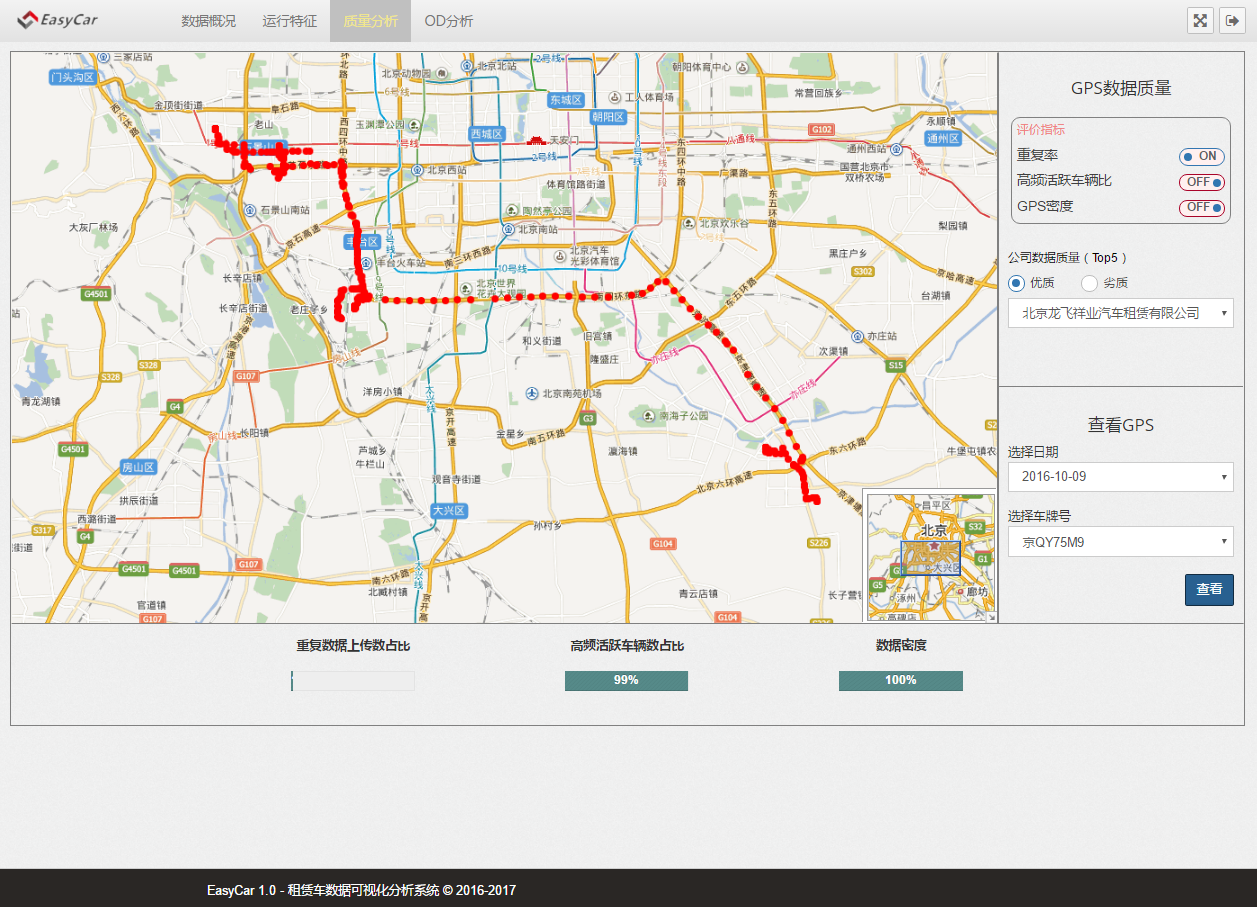
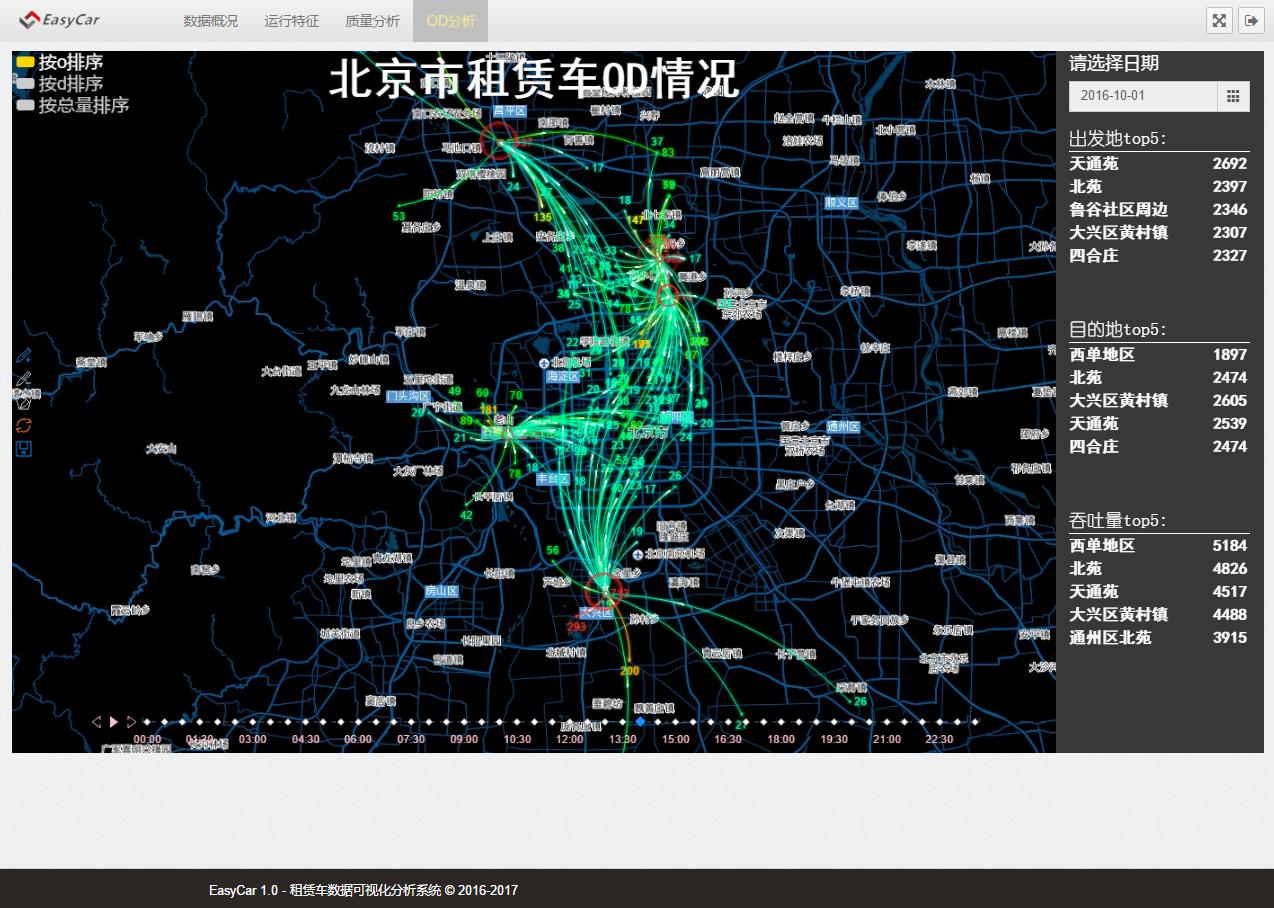


图.11 OD分析页面



## 本章小结

本章主要介绍了系统的整体设计。3.1节介绍了系统框架，将系统分为数据管理层，后台处理层以及前端展示层。3.2节介绍了数据库管理工具MySQL，并设计了租赁车主要数据的ER图。3.3节介绍了后台使用的框架SSM，后台的层次架构，并阐述了每个层次的功能。3.4节介绍了前端界面使用的可视化工具Echarts，以及系统四个界面的设计图。

# 系统实现

## 租赁行业的数据概况展示

租赁行业的数据概况包括三个部分：基础数据、三种租赁模式车辆占比情况、GPS数据量随时间变化曲线。

### 基础数据展示

该功能主要是利用立体柱状图展示基础数据在8-11月份的数据量，以及增减趋势。基础数据包括：车辆数、合同数、备案数。

1. 数据获取

基础数据比较简单，只需从每个月原始的租赁数据文件（CSV格式）中即可获得。租赁文件由车辆文件、租赁合同、备案文件组成。

1. 后台处理

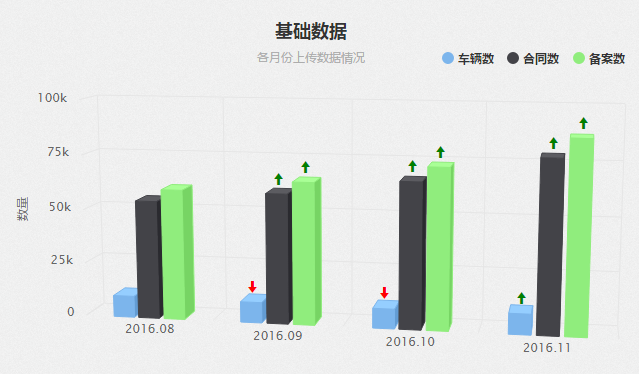
后台使用Java类File，对每个月的车辆数据、租赁合同数据、备案数据文件进行行数统计。将统计好的三个数据转换成JSON格式，提供给前端获取。数据格式如表4.1所示。

|  |  |  |
| --- | --- | --- |
| 表4.1 基础数据 | | |
| 名称 | 含义 | 样例 |
| name | 数据名 | “车辆数” |
| value | 4个月的统计值 | [9912, 9444, 9109, 9612] |

1. 前端展示

为了直观的掌握基础数据的整体情况，前端采用立体柱状图的形式进行展示。横坐标表示各月份，纵坐标表示数量，分别使用蓝黄黑三个条形图表示三种数据。为了了解各月份的数据增减趋势，还根据每个月的数据变化在相应的条形图上添加上升和下降箭头。效果如图4.1所示。

图.1 基础数据效果图



### 各租赁类型车辆占比情况

该功能主要为了展示每个月三种租赁类型的车辆占比情况。用户点击日期选取控件，选择想要查看的月份，前端通过后台提供的接口获取该月的车辆数据并展示。

1. 数据获取

由于租赁车表里面没有时间字段，为了获取不同月份的租赁车辆数，需要从GPS表里面进行统计。例如获取长租类型2016年8月份车辆数的SQL语句如下：

SELECT COUNT(DISTINCT plateNumber)  
FROM gps

WHERE date like “2016-08%” and type=1

因为GPS表中相同车有多个GPS数据，因此需要使用DISTINCT指定不同的车牌号。type的取值可以为1,2,3分别表示租赁类型为长租、短租和分时租赁。

1. 后台处理

Service层通过调用DAO层的接口，分别获取指定月份下的长租、短租和分时租赁的车辆数。Controller将Service层传上来的数据，加工处理成JSON格式，提供给前端获取。数据格式如表4.2所示。

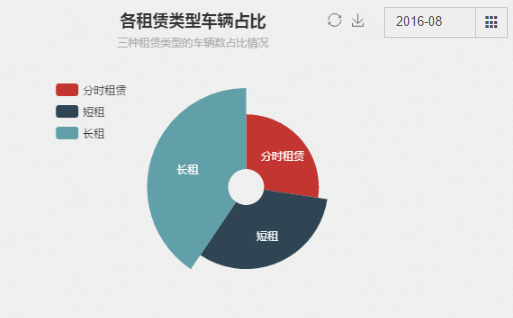
|  |  |  |
| --- | --- | --- |
| 表4.2 各类型车辆数 | | |
| 名称 | 含义 | 样例 |
| type | 租赁类型 | “长租” |
| value | 车辆数 | 987 |

1. 前端展示

为了体现三种租赁类型的占比情况，前端使用Echarts独具特色的南丁格尔玫瑰图（特殊饼图）。为了增加可交互性，系统还为玫瑰图的扇形增加了监听事件，用户可以拖曳重计算，将几种租赁类型进行合并。右上角为日期选取控件，用户点击该控件，可以选取想要了解的月份。效果如

图4.2所示。

图.2 各租赁类型车辆占比效果图



### GPS数据量随时间变化曲线

该功能主要展示三种租赁类型每天的GPS上传量，通过动态变化曲线的形式了解各类型车辆上传 GPS数据的情况。

1. 数据获取

这部分的数据为三种租赁类型每天的GPS数据量，因此可以从表gps获取该数据。例如获取分时租赁类型2016年10月30号的GPS量的SQL语句如下：

SELECT COUNT(\*)

FROM gps

WHERE type=3 and time like “2016-10-30%”

1. 后台处理

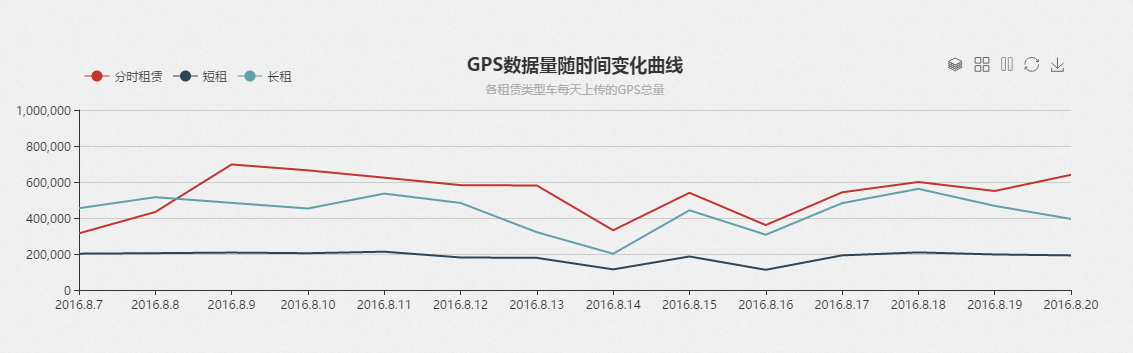
Service层循环调用DAO层的数据获取接口，将三种租赁类型的所有日期的数据量依次获取，保存在三个List中。Controller获取到List之后，将List转换成JSON格式，提供给前端获取。数据格式如表4.3所示。

|  |  |  |
| --- | --- | --- |
| 表4.3 GPS数据量 | | |
| 名称 | 含义 | 样例 |
| type | 租赁类型 | “分时租赁” |
| value | 所有日期的GPS数据量 | [32001,557242,…] |

1. 前端展示

为了体现GPS数据量随时间变化的曲线，前端采用Echarts的折线图，然后配合定时加载数据函数setInterval(loadData,1000)，每隔1000毫秒（1秒）刷新数据，达到GPS数据量按日期的动态变化效果。在右上角的toolbox中有暂停开始按钮，使用户可以随时暂停曲线的动态效果，以便查看某天的数据情况。效果如图4.3所示。

图.3 GPS数据量随时间变化曲线效果图



## 租赁车辆运行特征分析

租赁车辆的运行特征由四个指标构成：租赁时长、使用强度、日均行驶里程、出行时间。这一功能主要为统计功能，所使用到的数据均可从原始租赁数据获取，数据统计好之后直接转换成JSON格式供前端使用，因此并没有将数据存入数据库中。

### 租赁时长

对租赁合同进行时长的统计，将各类型换分成不同的时间段。长租按月进行划分成[2个月，3个月，4个月，5个月，6个月，12个月，>12个月]。短租按天进行划分成[1天，7天，14天，21天，>21天] 。 分时租赁按小时进行划分成[1小时，2小时，3小时，4小时，5小时，6小时，>6小时]。

1. 数据格式

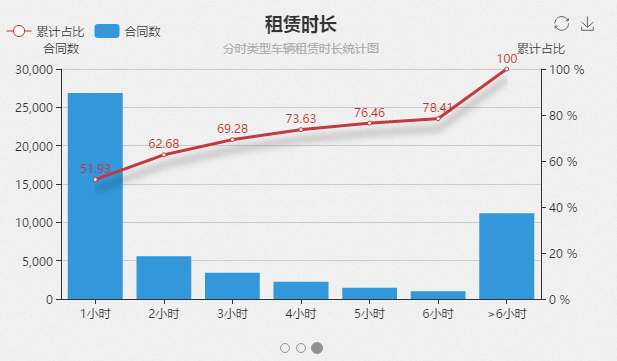
按照上述时间段的划分方式，从原始租赁数据分别统计各时间段的合同数。统计完成后，将结果转换成JSON格式。数据格式如表4.4所示。

|  |  |  |
| --- | --- | --- |
| 表4.4 租赁时长 | | |
| 名称 | 含义 | 样例 |
| type | 租赁类型 | “分时租赁” |
| value | 各时间段的合同数 | [26875,5562,3416,…] |

1. 前端展示

前端使用柱状图以及折线图组合的形式租赁时长特征。横坐标为租赁时长时间段，左纵坐标为合同数，右纵坐标为累计占比。使用蓝色条形图表示各时间段的合同数，红色折线图表示各时间段的合同数累计占比。图形下有三个圆点按钮，点击可查看相应租赁模式下的结果。鼠标在图形之外时，三个图会每隔5秒进行轮播，达到动态效果。效果如图4.4所示

图.4 租赁时长效果图



### 使用强度

使用强度特征表示的是租赁车每天的出行次数。根据对原始数据统计可将各租赁类型车辆的出行次数划分成[2次，4次，6次，8次，10次，>10次]。为了分析双休日与工作日的差异，还分别统计了两种日期的平均出行次数。

1. 数据格式

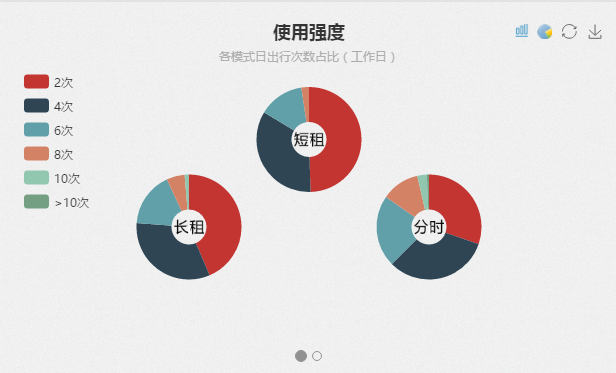
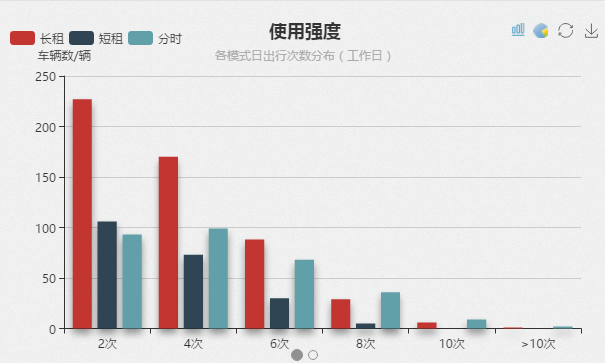
按照上述的划分方式，分别统计各租赁模式的出行次数。统计完成后，将结果转换成表4.5的格式。

|  |  |  |
| --- | --- | --- |
| 表4.5 使用强度 | | |
| 名称 | 含义 | 样例 |
| type | 租赁类型 | “长租” |
| isWorkday | 是否工作日，1表示工作日，0表示双休日 | 1 |
| value | 该租赁类型下各出行次数段的车辆数 | [227,170,88,29,6,1] |

1. 前端展示

首先使用柱状图展示三种租赁类型的出行次数。横坐标为日出行次数，纵坐标为车辆数。红黑蓝三种条形图分别表示各租赁类型的数据。点击图形下方的两个圆点，可以进行工作日和双休日的切换。为了更直观的观察出行次数的占比情况，在图形右上角添加了自定义饼图变化控件，用户点击该控件，柱状图就会变换成三个饼图。效果如图4.5所示。

图.5 使用强度效果图



### 日均行驶里程

该功能展示了三种租赁模式在工作日和双休日下的日均行驶里程。按照统计结果，将行驶距离划分为多个数据段，最短距离为10千米，最长为>500千米。

1. 数据格式

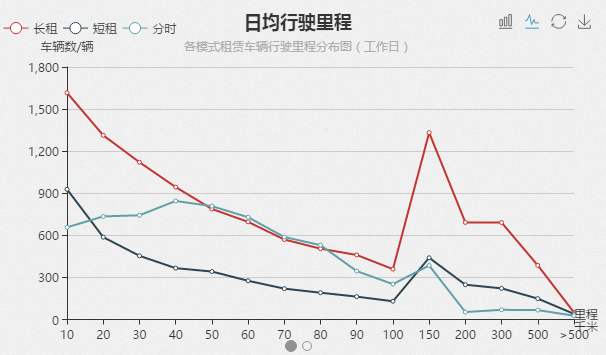
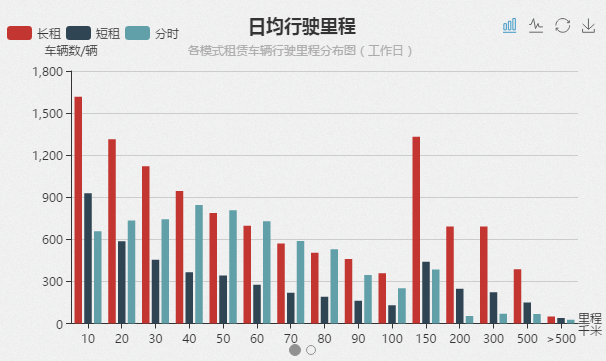
按照上述划分方式，将统计结果转换成如表4.6的格式。

|  |  |  |
| --- | --- | --- |
| 表4.6 使用强度 | | |
| 名称 | 含义 | 样例 |
| type | 租赁类型 | “短租” |
| isWorkday | 是否工作日，1表示工作日，0表示双休日 | 0 |
| value | 该租赁类型下各里程段的车辆数 | [245,170,156,…] |

1. 前端展示

前端采用柱状图和折线图两种形式展示结果。横坐标为里程数，纵坐标为车辆数。用不同的颜色区分三种租赁类型的数据。点击图形下方的圆点，可以进行工作日和双休日的切换。右上角有柱状图和折线图的小图标，点击图标可以进行图形间的转换。由于该功能数据较多，为了更清晰地查看局部数据，在图形内增加了放大缩小功能。整体效果如图4.6所示。

图.6 日均行驶里程效果图



### 出行时间

出行时间特征展示的是租赁车在一天中各个时间段的出行次数。将一天按照小时划分成24个时间段，统计各时间段内的出行次数，也分别统计了工作日和双休日的出行情况。

1. 数据格式

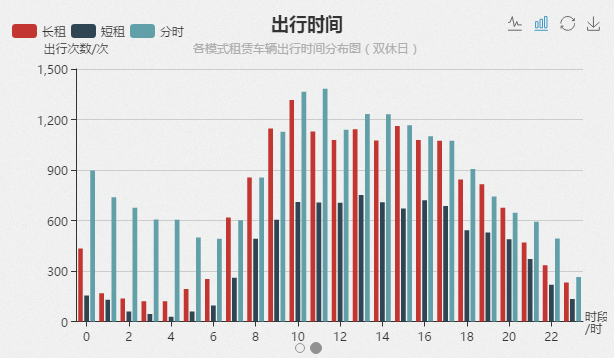
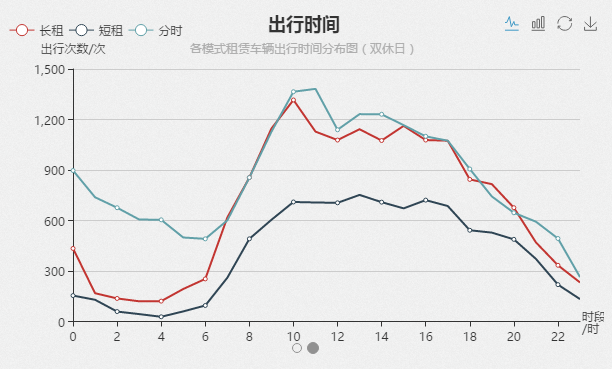
按照上述的时间段划分方式，将统计结果转换成如表4.7的格式。

|  |  |  |
| --- | --- | --- |
| 表4.7 使用强度 | | |
| 名称 | 含义 | 样例 |
| type | 租赁类型 | “分时租赁” |
| isWorkday | 是否工作日，1表示工作日，0表示双休日 | 1 |
| value | 该租赁类型下各时间段的出行次数 | [2567,2194,2011,…] |

1. 前端展示

前端采用折线图和柱状图两种形式展示结果。横坐标为时间段，纵坐标为出行次数。用不同颜色区分三种租赁类型。点击图形下方的圆点，可以进行工作日和双休日的切换。点击右上角的折线图和柱状图的小图标，可以进行图形的转换。效果如图4.7所示

图.7 出行时间效果图



## 各公司GPS数据质量分析

该部分主要分为两个功能：第一，根据三个评价指标，对所有公司的GPS数据质量进行评价从而评选出TOP5的优质公司和TOP5的劣质公司；第二，查看某辆租赁车一天的行驶轨迹。

### 公司评比

首先统计各公司上传的GPS总数。由于数据库表gps（结构如表4.8所示）中没有公司字段，因此需要结合表vehicle（结构如表4.9所示）进行统计。每条GPS记录，都有车牌号，通过车牌号可以在表vehicle中找到对应的公司ID，该公司的GPS总数加1。如此，就可以统计所有公司的GPS总量。

|  |  |  |
| --- | --- | --- |
| 表4.8 数据库表gps | | |
| 名称 | 数据类型 | 说明 |
| id | INT | 主键 |
| time | VARCHAR(20) | gps上传时间 |
| plateNumber | VARCHAR(45) | 车牌号 |
| type | INT | 租赁类型，1,2,3分别表示长租，短租，分时租赁 |
| lng | DOUBLE | 经度 |
| lat | DOUBLE | 纬度 |

|  |  |  |
| --- | --- | --- |
| 表4.9 数据库表vehicle | | |
| 名称 | 数据类型 | 说明 |
| id | INT | 主键 |
| plateNumber | VARCHAR(45) | 车牌号 |
| type | INT | 租赁类型，1,2,3分别表示长租，短租，分时租赁 |
| comId | INT | 公司ID |

1. 评价指标
2. 重复数据上传占比

很多公司的GPS数据都存在这冗余数据过多的问题，虽然上传的量足够大，但有着较多的重复数据，重复数据占比越大，证明重复上传率越高。

|  |  |
| --- | --- |
|  | （4.1） |

由公式4.1可知只需统计公司上传的重复GPS数据量，然后除以该公司的GPS总量，就是重复上传数据占比。

1. 高频活跃车辆占比

将有GPS数据上传的车辆称为活跃车，将GPS上传时间间隔小于2分钟的车辆定义为高频活跃车辆，单天中，高频活跃车辆越多、占比越大，其车辆轨迹的重塑性越高。

|  |  |
| --- | --- |
|  | （4.2） |

根据数据库表gps的time字段，就可以统计一天当中某辆车gps上传的平均间隔时间，如果该时间小于2分钟，则该车对应公司的高频活跃车辆加1。通过表gps和表vehicle就可以知道上传了GPS数据的车辆有哪些，由此可以计算每个公司的活跃车总量。由公式4.2可知，高频活跃车数量除以活跃车总量即为高频活跃车辆占比。

1. 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。计算两个GPS坐标点之间的距离见式4.3。GPS密度越大，GPS质量越好。GPS密度公式见式4.4

|  |  |
| --- | --- |
|  | （4.3） |

其中Lat1 Lng1表示A点经纬度，Lat2 Lng2表示B点经纬度；

a=Lat1-Lat2 b=Lng1-Lng2

6378.137位地球半径，单位是千米。

|  |  |
| --- | --- |
|  | （4.4） |

在表gps中，依据公式4.3对同一辆车相邻的2个GPS数据进行距离计算，如果距离小于1000米，则该公司的高密度GPS数量加1。最后根据公式4.4，高密度GPS数量除以GPS总数，即为该公司的GPS密度。

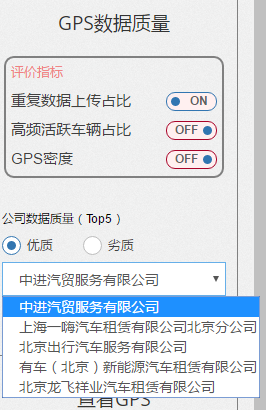
将上述计算的三个指标数据存入数据库表company\_indicator（公司指标）中，结构如表4.10所示。

|  |  |  |
| --- | --- | --- |
| 表4.10 数据库表company\_indicator | | |
| 名称 | 数据类型 | 说明 |
| comId | INT | 公司ID |
| repeatRate | FLOAT | 重复数据上传占比 |
| highActiveRate | FLOAT | 高频活跃车辆占比 |
| gpsDensity | FLOAT | GPS密度 |

1. 处理过程

前端展示效果如图4.8所示

图.8 公司评比效果图



当用户点击某个评价指标开关时，表示以该指标对公司进行评比，评价指标开关可以同时开多个，即使用多个指标评比。

如果只用一个指标进行评比，则只需要在表company\_indicator中使用order by SQL语句即可（注意，重复上传比越小质量越好，因此只需升序排列，而其余两个指标是值越大质量越好，因此需要使用DESC降序排列），然后再根据公司ID就能找到对应公司名。

如果使用多个指标评比，后台需要对多个指标得到的多个结果进行加权评分，对每个公司进行评分，然后再按照分数的高低进行排序。

用户还可以点击“优质”“劣质”单选框，查看数据质量优（或劣）的TOP5公司。

在界面的底部，还用条形图展示选中公司的三个指标值，效果如图4.9所示。

图.9 三个指标值

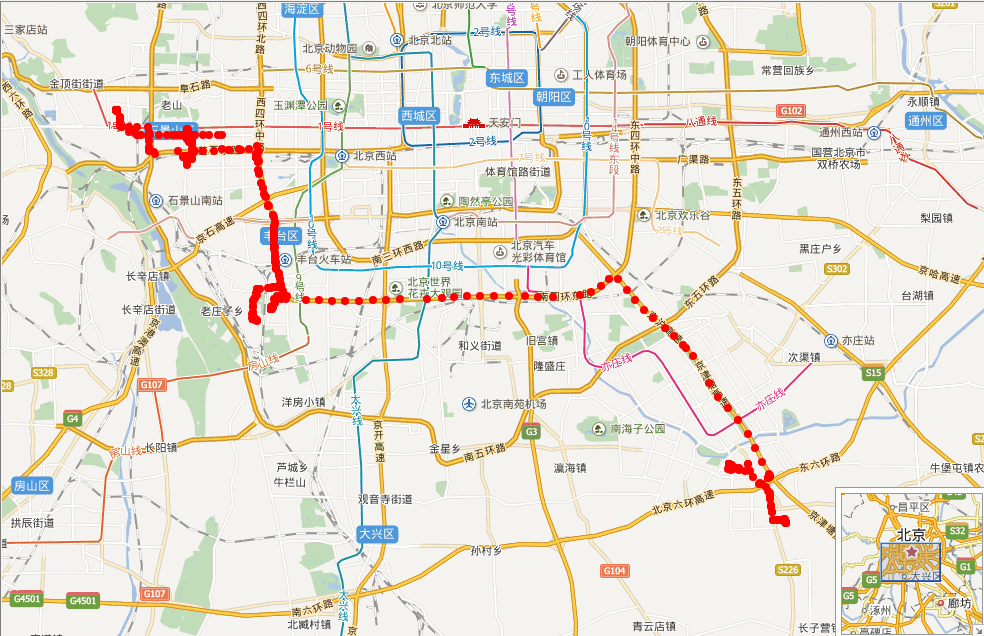


### GPS轨迹展示

为了了解公司上传GPS的具体情况，本系统还添加了查看某辆车某天的GPS轨迹功能。

用户首先选取某个公司，后台根据公司名找到属于该公司的所有车辆的车牌号展示在前端。然后，用户再选取日期和车牌号，后台将获取这天该车辆所有的GPS。由于GPS轨迹的展示使用百度地图，因此需要调用百度地图API的接口将GPS的经纬度转换成百度坐标。转换完之后，使用百度地图添加覆盖物的功能（addOverlay），在地图上用红色小圆点表示GPS轨迹。前端效果如图4.10所示

图4.10 GPS轨迹图



## 租赁车OD情况分析

# 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

附录