



東南大學  
SOUTHEAST UNIVERSITY

# 软件项目管理与实践

BSI 软件项目小组

东南大学计算机科学与工程学院、软件学院

School of Computer Science & Engineering College  
of Software Engineering

Southeast University

二 0 一 九 年 四 月

## 小组成员与角色

学号	姓名	角色
71116111	刘骁	项目经理
71116302	马欣宇	产品经理
71116233	白丰硕	质量经理
71116236	叶绵和	系统架构师
71116216	杨昱昊	前端开发工程师
71116218	杨志成	后端开发工程师
71Y16111	陈泽远	硬件开发工程师
71116114	李寒松	测试工程师

# 一、环境背景

随着机动化水平的不断提高,城市道路交通安全形势日趋严峻。近年来,机动车快速增长、人口急剧增加,道路基础设施建设已无法及时满足人们的交通出行需求。同时,大量的机动车也容易造成交通拥堵、路网瘫痪、事故频发等问题,其实交通安全,正面临日益严峻的压力和挑战。如何减少事故发生起数、减轻事故严重程度,合理规划道路交通已成为人们普遍关注的问题之一。结合道路交通事故随机事件的特点,利用多元统计方法分析事故及相关因素数据,可有效获取风险因素对事故影响的显著规律,从而提出有效改善措施。

在图 1.1 中可以看出,洛杉矶近年来的事故发生趋势呈明显的上升,事故造成的死亡人数也急剧增长。

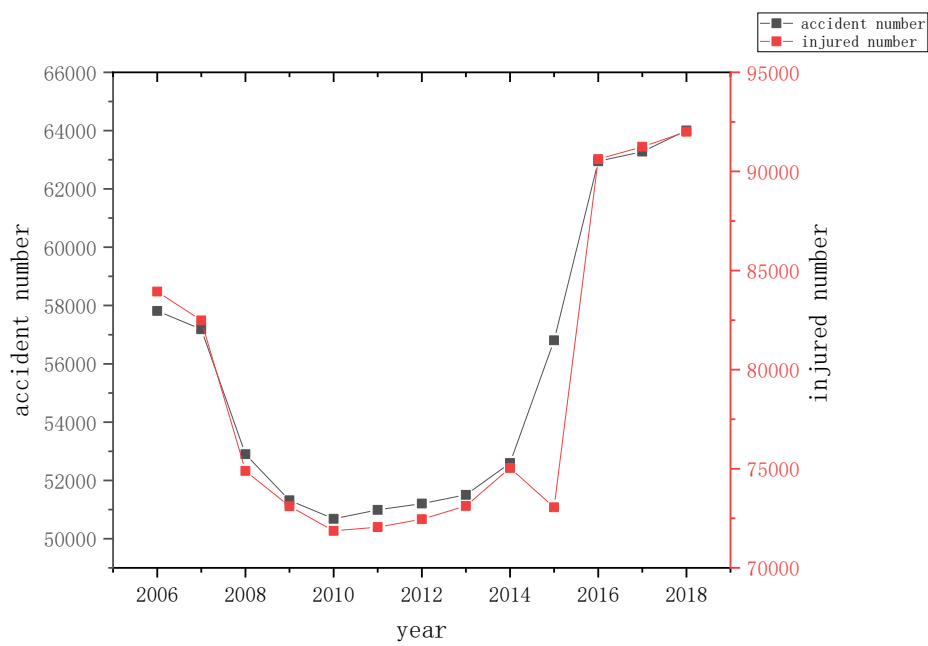


Fig 1.1

在一段路网中,基于统计信息可以发现某些路段或者交叉口频繁发生事故,这些路段或交叉口通常称为黑点。如图 1.2 所示,在对洛杉矶县的事故信息统计后,可以发现在地图中标为红色的三个点是事故高发的路段(黑点)。

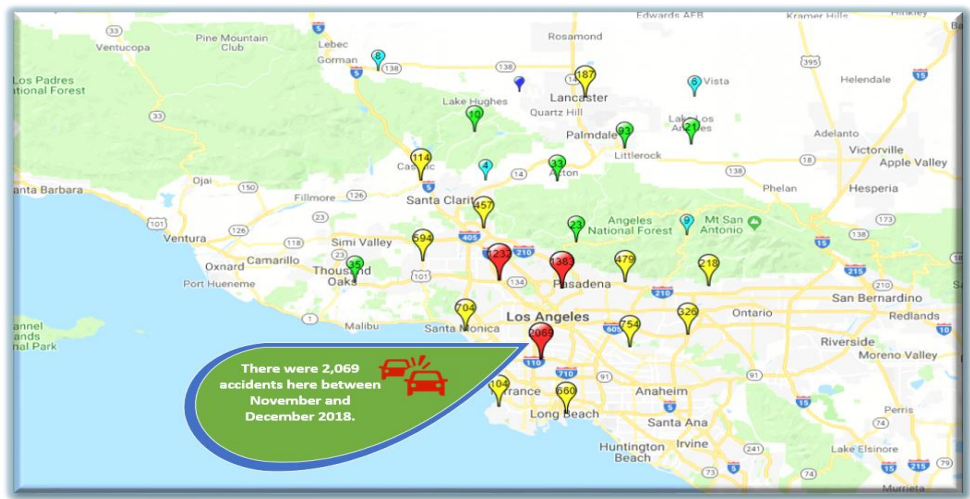


Fig 1.2

我们能够提出一个模型，来科学有效地识别这些黑点，就能帮助决策者制定更加有效的道路交通安全举措，同时也能在导航中提醒驾驶者黑点的位置，并且规划更加科学的路线，以降低事故的发生率。然而，上图中仅仅以一段时间内事故的数量来判别黑点，这是不全面的，因为它不能充分地利用可获取的多源数据，例如道路结构、道路亮度、人流量、天气情况等。同时，交通事故的发生与时间和空间也存在一定关系。我们小组将要开发出一款综合多因素的路网黑点识别系统，就能提供更加准确全面的黑点识别结果，以供驾驶人员和决策者参考，以此来满足人们对道路交通安全的需要。

另一方面，道路交通以及事故数据信息的分析可用于改善道路和交通工程设计,为政府部门实施交通安全管理提供决策支持,还有助于针对机动车驾驶员设计交通安全教育项目,提高驾驶员的安全意识,从而改善城市道路交通安全状况。目前我国城市交通发展的历程相当于西方国家的 60~70 年代,与发达国家相比,城市机动车密度还比较低。尽管如此,由机动车引发的环境污染问题和城市交通堵塞问题也很严重。这充分体现了我国在道路规划和建设方面还存在体制上、行政上和技术上的问题。我们考虑更多的是从技术层面来解决问题。城市网络很复杂,交通的运行很复杂,产生交通问题的因素也很复杂,相应制定的城市交通管理方案往往由多个管理策略、管理措施组合而成,任何一个建设或管理措施的实施都会引起整个城市路网上交通运行情况的改变。并且需要在过往的建设规划中,更多倾向于城市之间的缓解城市物流交通压力,促进经济发展等现实因素。在高新技术发展的今天,也要将以人为本的思想贯彻下去,要将交通事故的因素也要考虑在内。

为此,需要明确城市道路的突出交通安全问题,利用多项事故统计指标对城市道路交通事故在不同时间、不同道路类型、不同路口路段类型、不同车辆使用性质以及不同事故类型上的分布特征进行了描述性统计分析,发现主干路和次干路、四叉信号交叉口和无控制出入路段、出租和公交客运的交通安全问题,以及机动车与自行车碰撞事故值得关注。在总结国内外道路交通事故分析研究的基础上,提出了城市道路交通事故分析概念模型,并结合上述交通安全突出问题,设计了城市道路交通事故分析体系,具体包括三个方面:城市干道网络中信号交叉口和无控制出入路段事故发生频次分析、城市道路机动车与自行车碰撞事故分析以及城市道路公共运输驾驶员“心理-行为-事故”影响链分析。

## 二、项目环境

### 开发环境:

首先利用 python 中 Scrapy 的爬虫框架进行数据采集。

项目数据达到 TB 级别,因此利用分布式系统 HDFS 存储数据。而对数据的处理,因为涉及到利用机器学习算法进行数据分析建模,采用 Spark 框架对数据进行处理。为了也能适用简单的数据处理,我们也使用 Hive 进行简单的数据清洗。

对于处理好的数据,使用 SpringBoot 框架,采用 SSM 架构,进行逻辑业务处理。最后前端使用 Echarts 进行数据展示。

各个版本号如下:

Python: 3.5

Scrapy: 0.22.2

Hadoop: 2.6  
Spark: 2.1.0  
Hive: 1..1.0  
SpringBoot: 2.0  
Java: 1.8  
Mabits: 3.0

## 测试环境:

测试环境需求取决于要测试的应用程序的类型。对于大数据测试,测试环境应包括:

- 它应该有足够的空间用于存储和处理大量的数据
- 它应该具有分布式节点和数据的集群
- 它应该具有最低的 CPU 和内存利用率,以保持性能高

### 功能测试步骤:

- a. 数据导入/预处理验证阶段
- b. MapReduce 数据输出验证阶段
- c. 验证大数据 ETL 到数据仓库
- d. 验证分析报告

### 非功能测试:

- a. 性能测试
- b. 容错性测试
- c. 可用性测试
- c. 扩展性测试
- d. 稳定性测试
- e. 部署方式测试
- f. 数据一致性测试
- g. 压力测试

## 生产环境:

假设现在生产环境的信息如下:

- 服务器数量: 6
- 操作系统: Centos7
- Master 节点数: 2
- Slave 节点数: 4

划分各个机器的角色如下:

主机名	角色	运行进程
hadoop1	Master	Namenode
hadoop2	Master_backup	Namenode
hadoop3	Slave、Yarn	Datanode、ResourceManager、NodeManager
hadoop4	Slave	Datanode、NodeManager
hadoop5	Slave	Datanode、NodeManager
hadoop6	Slave	Datanode、NodeManager

### 三、项目组成

我们的项目主要由 Web 模块、大数据处理模块和数据信息采集模块组成。Web 模块中主要使用到了 SpringBoot、MySQL、Vue.js 和 Single Page Application 等技术。大数据处理模块中使用了 Hadoop、Spark、Spark streaming 和 Storm 等技术。

#### 数据信息采集模块

##### 环形感应线圈

该设备通常由环型线圈、传输馈线、检测处理单元部分组成。环形线圈铺设在道路上，并在其附近形成一个磁场，当车辆进入这个磁场时，检测处理单元就检测到一个车辆，并输出信号。它不仅可以计数、检测交通流量，还可以测速。

##### 微波检测器

该设备是一种雷达探测器，利用多普勒效应检测车辆，不仅可以检测交通量，还可以测速，从而达到检测道路交通信息的目的。

##### 视频监控系统

该设备主要由前端、传输和终端三大部分组成。前端部分：主要包括摄像机、镜头、云台、解码器等；传输部分：常用的有光缆、视频电缆、电话线等；终端部分：监视器，它可以显示前端传来的图像，还可对前端设备进行控制。电视监控系统有图像监视和交通数据采集双重功能；安装简单，检测率高，寿命长，维护费用低。

##### 浮动车交通信息采集

这种方式是通过安装有全球卫星定位系统和无线通信装置的车辆（如出租车、公交车等）与交通数据中心进行实时信息交换，特点是采集范围广、投资少，能够反映路网运行状态的变化，为疏堵提供参考。

## Web 部分

### SpringBoot

Spring Boot 是由 Pivotal 团队提供的全新框架，其设计目的是用来简化新 Spring 应用的初始搭建以及开发过程。该框架使用了特定的方式来进行配置，从而使开发人员不再需要定义样板化的配置。通过这种方式，Spring Boot 致力于在蓬勃发展的快速应用开发领域(rapid application development)成为领导者。其优点是简单，快速，方便，其特性有：

- 能够创建独立的 Spring 应用
- 本身嵌入了 Tomcat、Jetty 容器
- 提供可选的 starter 依赖库简化应用构建配置
- 自动配置 Spring 以及第三方依赖
- 提供生产级的特性，如度量、检查和外部化配置
- 无代码生成并且不需要 XML 配置

### MySql

MySQL 是一个关系型数据库管理系统，瑞典 MySQL AB 公司开发，目前属于 Oracle 旗下产品。MySQL 是最流行的关系型数据库管理系统之一，在 WEB 应用方面，MySQL 是最好的 RDBMS(Relational Database Management System，关系数据库管理系统) 应用软件。

MySQL 是一种关系数据库管理系统，关系数据库将数据保存在不同的表中，而不是将所有数据放在一个大仓库内，这样就增加了速度并提高了灵活性。

MySQL 所使用的 SQL 语言是用于访问数据库的最常用标准化语言。MySQL 软件采用了双授权政策，分为社区版和商业版，由于其体积小、速度快、总体拥有成本低，尤其是开放源码这一特点，一般中小型网站的开发都选择 MySQL 作为网站数据库。

由于其社区版的性能卓越，搭配 PHP 和 Apache 可组成良好的开发环境。

### Vue.js

Vue.js 是一套构建用户界面的渐进式框架。与其他重量级框架不同的是，Vue 采用自底向上增量开发的设计。Vue 的核心库只关注视图层，并且非常容易学习，非常容易与其它库或已有项目整合。另一方面，Vue 完全有能力驱动采用单文件组件和 Vue 生态系统支持的库开发的复杂单页应用。

Vue.js 的目标是通过尽可能简单的 API 实现响应的数据绑定和组合的视图组件。

Vue.js 自身不是一个全能框架——它只聚焦于视图层。因此它非常容易学习，非常容易与其它库或已有项目整合。另一方面，在与相关工具和支持库一起使用时，Vue.js 也能完美地驱动复杂的单页应用。

## Single Page Application

单页 Web 应用 (single page web application, SPA), 就是只有一张 Web 页面的应用。单页应用程序 (SPA) 是加载单个 HTML 页面并在用户与应用程序交互时动态更新该页面的 Web 应用程序。浏览器一开始会加载必需的 HTML、CSS 和 JavaScript, 所有的操作都在这张页面上完成, 都由 JavaScript 来控制。因此, 对单页应用来说模块化的开发和设计显得相当重要。

## 大数据部分

此部分形象化描述目前流行的的大数据处理框架。

### Hadoop

Hadoop 核心组件为 HDFS 与 MapReduce。

其中 HDFS 是作为大数据存储标准, 亦是文件管理方式。是目前公认的高效的分布式管理系统。一种分布式的计算方式指定一个 Map 函数, 用来把一组键值对映射成一组新的键值对, 指定并发的 Reduce 函数, 用来保证所有映射的键值对中的每一个共享相同的键组。简述其过程则是数据传入 map, 处理后 reduce 将对数据进行整合、过滤或者合并等。

### Spark

spark 是批处理框架。其出现是基于提高 Hadoop 的效率, 在 Hadoop 基础上进行的提高。其核心仍是 Hadoop 技术。其能实现短时要求的处理。

### Spark streaming

Spark streaming 是一种伪实时流处理框架, 因为其仍是 spark 底层, 只是加了时间片, 在时间片内的数据会在时间抵达后进行处理, 给用户伪实施流处理的感觉。

### Storm

实时流处理框架。其核心抽象来讲, 数据像是乘客, 处理数据的部分像是车厢。列车向前, 每到一个站台数据进行上下车, 就像是乘客上下车厢。除非车停了, 否则数据就实时在流动处理。



## 四、项目目的

本项目是一个基于道路路段信息以及此路段发生事故的时间地点信息，从而对当前道路网络上的事故高发点进行计算的系统。对事故发生的时间、地点、天气、死伤人数等诸多数据进行清洗、整理、汇总、挖掘，从而构建出一个能够对已知、未知路段的事故高发地点进行自动化科学预测的模型，同时通过这样的模型，我们也可以对未建设路段进行预测，判断待预测道路建设方案的事故概率是否可以接受。

基于以上两点核心能力，本项目中将构建一套中间件用来进行将大数据处理和具体业务分离。大数据处理过程包括对现有道路事故数据的筛选处理，以及建立一个用来预测的模型。从而适配多种具体业务，主要业务包括以下几点：

1. 针对个人用户推出带有鲜明特色的导航系统，与其他系统不同，我们的导航系统加入了对风险地带的考量。在面对事故高发点时可以对用户作出提示，并且为用户提出其他可供选择的方案来提升用户的安全，
2. 针对于城市的道路规划，我们也推出道路风险预测系统，面向城市道路规划工作者。用户只要将地区内的道路信息、事故信息和待建设道路信息输入系统，就可以通过待建设道路上是否会出现事故高发点来对建设方案进行考量，为城市的规划提供可参考元素。
3. 针对于交通部门，我们将向交管部门展示道路风险模型对事故高发点的预测，帮助交通管理部门进行后续的监管和道路设施的改进工作。
4. 同时，我们可以通过中间件构建公有的 API 服务，这些 API 可以用户商用项目或者个人项目，搭建一个 BSI 路网黑点开放平台。我们可以通过授权费用来获得收益，同时让更多用户享受到我们的服务。