**区域客流预警监控系统的设计与实现**

倪嘉琦

(上海交通大学)

**摘要：**防范类似“2014年底上海外滩踩踏事件”、“2015年麦加朝觐踩踏”等公共安全事故的发生是各国政府部门非常重视的事情。因此，利用有关部门提供的多维度数据和成熟的时间序列预测算法，可以实现对城市人流聚集情况进行分析与预警，从而为政府公安机构提供事前预防和事后分析的支持并可保障市民出行的安全。该系统把市民出行的行为与城市地理区域位置相结合，以客流监控、区域人口分析为出发点，利用交通卡出入站闸机AFC系统以及电信OIDD信令数据，经过融合、汇总和分析为该系统提供了有力的数据支撑。系统测试以及六个月的实际上线运行表明，该系统是可行的和有效的。

**关键词：**客流监控，人群密度预警，客源流动分析，时间序列预测

**Design and implementation of regional passenger flow early warning and monitoring system**

Ni Jiaqi

(SJTU,SITI, SSC,Shanghai,China)

**Abstract：**To prevent " 2014 Shanghai Bund stampede", "2015 Hajj stampede" public safety accident occurs as governments around the world attach great importance to the things. Therefore, the use of existing data and large data technology, can achieve the analysis and early warning of urban traffic flow, to protect the safety of the public, and to provide support for the government's public security agencies in advance prevention and ex post analysis. The system combines the behavior of citizens' travel with the location of the city, and takes the passenger flow monitoring, the regional population analysis as the starting point, and uses the AFC system of traffic card and the OIDD data, which provides a powerful data support for the system.

**Key words:** Traffic monitoring, population density warning, tourist flow analysis, time series model

# 1 引论[[1]](#footnote-1)

目前，随着客流预警监控正逐步进入“大数据时代”，客流监控越来越受到有关部门的重视**[1]**。利用大数据技术预测手段可以防范类似“2014年上海外滩踩踏”等事件的发生，某研究院提出，利用有关部门提供的已有多维度的数据和成熟的时间序列预测方法，设计并实现区域人流预警监控系统，该系统可实时获取特定区域内人员数量，当预测客流数量激增或异常时，可做到自动告警，达到实时客流量监控预警的目的。

针对上述需求，本文将以实习单位(C研究院)的实际项目为背景，设计并实现一个区域客流预测系统。短期的区域客流预测是处置与管理大客流突发事件的重要方法，相对于大范围长期的预测而言难度更大。过去，在城市建设与管理的初期阶段，由于调查数据少，统计方法简单，对区域客流预测尚未放到重要位置，缺乏系统认识。而现在，随着大数据时代的到来，区域客流预测越来越受到有关部门的重视**[2]**。因此，为了保障市民出行安全以及城市管理成本，该系统的开发与实现是有重大意义的。本文结合差分自回归移动平均模型设计了一套基于预测的客流监控系统，并对其进行了实现，最后对该设计的系统进行效果分析与总结。

# 2客流预测方法

常用的短时客流预测方法主要有以下三种模型：基于统计学理论模型、基于非线性理论模型和基于神经网络模型**[3]**，具体方法如图2-1所示：



图2-1 短时客流预测方法

Fig.2-1 The forecasting methods of passenger flow on short-term

上述预测模型都有其相适应的场景，其中差分自回归移动平均模型(Autoregressive Integrated Moving Average Model,简记ARIMA)是一种广泛应用于短期预测的时间序列模型，是由博克思(Box)和詹金斯(Jenkins)于70年代初提出的一著名时间序列预测方法。这个模型一旦被建立后就可以从时间序列的过去值及现在值来预测未来值。其中，AR是自回归，p为自回归项；MA为移动平均，q为移动平均项数，d为时间序列成为平稳时所做的差分次数。

时间序列分析模型的基本思想是用已有的数据序列预测未来。在时间序列分析中，数据的属性值是随着时间不断变化的。时间变化主要受到长期趋势、季节变动、周期变动和不规则变动四个因素的影响。SARIMA模型适用于具有季节周期性的时间序列分析，因此该方法适合本文研究的区域客流短期预测。时间序列模型建模简单，在历史数据充分的条件下可以获得较高的预测精度。

本文采用的数据选取了具有早晚高峰代表性的星期一，对它进行时间序列的平稳性训练，由图2-1可以得到的是上海人民广场2015年1月至5月每周一的客流曲线，从中可以看出客流呈现周期为7的变化规律，而由于2月16日接近春节，所以客流有一个明显的下降，约为往常的一半，同样地，4月6日是清明节放假，所以也有一个明显的客流低峰，其他时间均有明显的早晚高峰特征。

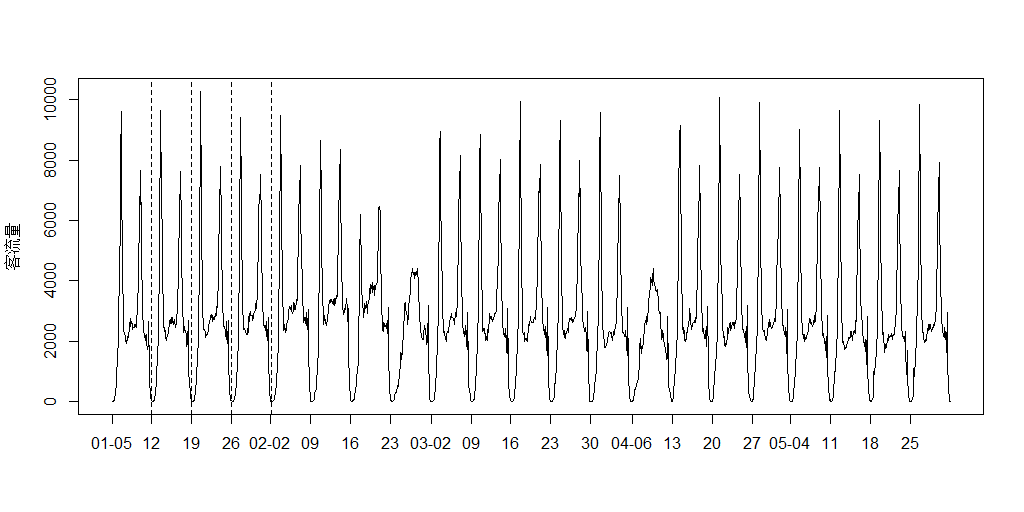


图2-2 2015年上半年人民广场客流曲线图

Fig.2-2 Passenger flow chart

由于源数据存在明显的不平稳性，所以需要对原始数据做一阶差分，由2-2左半部分可知，未能完全消除不平稳性，于是再做二阶差分，可以发现图2-2右半部分均值几乎为0，由此可知，随着时间推移，时间序列大致保持不变，因此设置差分项d=2较为合适。

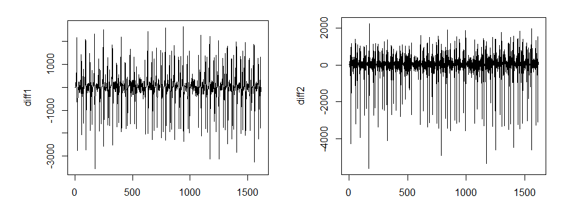


图2-2 一阶和二阶差分图

Fig.2-2 First order and two order difference graphs

接下来，需要选择合适的ARIMA模型参数**[4]**，即确定ARIMA(p,d,q)中合适的 p、q 值，通过R语言xts包中的“acf()”和“pacf()”函数来做判断，该序列带有明显的周期性和季节性变化。

auto.arima给出了最佳建议模型参数为ARIMA(3,0,3)(0,0,2)[7]。

ACF检验说明残差没有明显的自相关性，Ljung-Box测试显示所有的P-value>0.5，说明残差为白噪声，因此模型检验通过，并利用该模型对下一阶段数据进行预测。

# 3总体设计

## 3.1数据计算模块设计

客流预测系统数据来源于出入站闸机AFC系统以及电信OIDD信令数据。计算环境采用20个节点的hadoop集群，本系统的部署结构如下图：



图3-1系统总体物理部署

Fig.3-1 System overall physical deployment

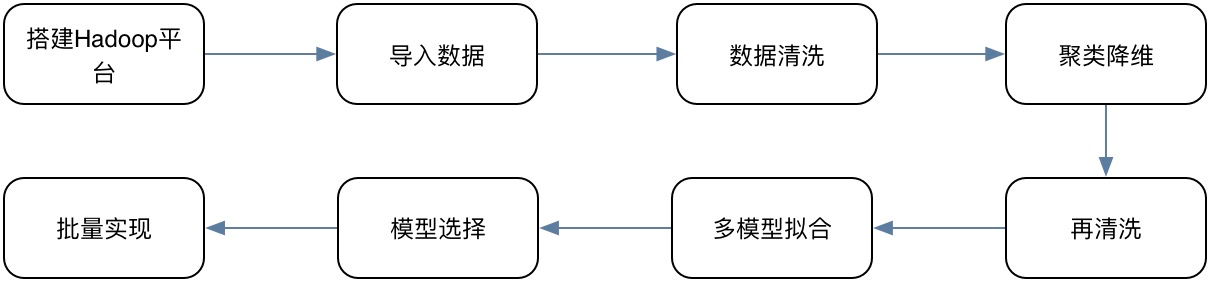
有限的信息需要通过海量的数据汇集来体现，因而对数据清洗的要求很高，数据清洗与处理的示意图如下：

图3-2 数据处理过程

Fig.3-2 Data processing procedure

系统应用架构图如下：

图3-3 系统应用架构图

Fig.3-3 System application architecture diagram

数据通过 sqoop 从业务数据库中抽取，使用 Hadoop + Hive **[11]**对数据进行抽取（extract）、转换（transform）、加载（load），然后将统计分析的结果再通过 sqoop 导入分析数据库中，通过 web service 进行数据展现以及提供数据接口。

## 3.2主要功能模块设计

预警模块分为两个部分：实时监控与客流预警。

客流实时预测模块主要通过对历史客流数据的分析和归纳，总结客流数据的趋势规律，并将其通过时间序列算法模型计算出未来可能的客流数据，区域客流预测的计数单位为15分钟。

事后分析模块分为两个部分：人群密度与地铁车站的进出客流查询。

人群密度：以热力图的形式展示24小时区域人数统计。

地铁车站的进出客流查询：进出站OD 用于展示轨道交通定制站点的全天进站客流以及OD流向情况，并列举TOP15 的指定站到其他站OD客流以及总计人数。

## 3.3接口设计

请求参数是外部系统在与平台进行数据交互时，提供给平台的请求数据，以便平台根据这些数据进一步处理。返回参数是平台根据业务请求，将处理结果返回给数据请求方。接口以Restful API**[6]**方式提供。客流预测系统采用RESTful Web Service 的整体设计风格，它是一个使用 HTTP 并符合 REST 原则的 Web 服务。通过 URL 可以传送 GET 请求，在表单指定 method="GET|POST" 来送出请求。在 REST 中的资源(Resource)代表整个网络上的资源。数据提供方提供各式各样的数据资源，而这些服务将由 URI来提供。处理过程有以下五步：

1）构造请求数据

外部系统根据交通卡数据服务平台（以下简称平台）提供的接口规则，通过程序生成需要传输给平台的数据集合。

2）发送请求数据

把构造完成的数据集合，通过页面链接跳转或表单提交的方式传递给平台系统。

3）平台系统对请求数据进行处理

平台系统得到这些集合后，会先进行安全校验等验证，一系列验证通过后便会处理完成这次发送过来的数据请求。

4）返回处理的结果数据

对于处理完成的交易，平台会将结果数据以JSON的形式反馈给数据请求方。

5）对获取的返回结果数据进行处理

数据交互图如下：

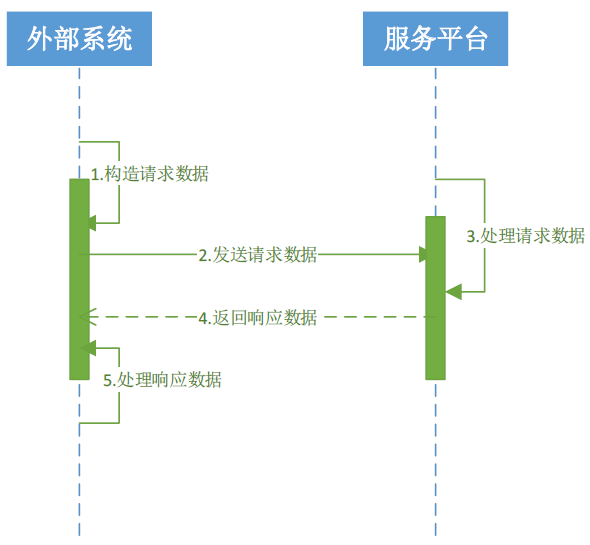


图3-4数据交互图

Fig.3-4 Data interaction graph

例如：

请求：

http://localhost:8080/xxxx

date =20150302

station ID =0111

应答：

{"result ":"00","inNum ":"xxxx "," outNum ":"xxxx "}

# 4区域客流预警监控系统的实现

本节将结合第二部分预测模型分析和第三部分的调度流程总体设计，对预测调度模型进行系统上的实现。结构图如下所示：



图4-1客流预警系统结构图

Fig.4-1 Passenger flow early warning system structure diagram

人员流动运行状态数据是进行大客流监控预警的核心数据。数据源主要包括：OIDD信令数据和交通卡刷卡数据。人员流动运行状态数据处理主要解决的问题是海量数据的实时处理与分析。人员流动运行状态数据处理的主要工作包括：

（1）基本变量计算**[5]**：核心是统计和计算各个采集点的人流的速度（加速度）、密度和流量，人员流动可表达为速度、流量、密度三者之间的关系：Q=k\*v

公式中Q 为客流的流率，k 为人流密度，v 为人流速度。

客流密度和客流速度主要依靠OIDD用户手机基站的切换数量、频率进行计算。

（2）客流的区域分布统计：将海量的用户基站信令数据，按照区域、时间段、移动方向进行分类和统计，计算出区域网格内的人员流入量、流出量、存量与密度，以及人员流动的方向和速度。

（3）基站与区域电子围栏数据处理：由于基站的数量是不断调整的，监控区域也是不断调整的，因此需要对基站的地理位置信息、以及与区域电子围栏数据的对应关系进行不断的维护。调整的数据需要及时反映到预处理程序当中，并更新到GIS图层数据当中。

以下是接口标准：

/trajectoryAnalysis/hourPeopleNum?area=区域名称&date=日期&hour=小时

数据需求者需要输入：

区域名称拼音：renminguangchang

日期格式如：2015-01-01

小时格式如：10

输入参数后可以正确返回该区域（人民广场2015-01-01上午10点到11点）的人数。

目前，随着城市的发展，市民出行的需求在逐渐增加，尤其客流是在开放性的、不同空间之间的流动，出行过程中具有很大的不确定性和不可预见性，而实时获取客流数据会极大提高出行效率与出行体验**[7]**。

# 5系统测试及应用

本章介绍系统测试与应用，系统测试包括功能性测试与非功能性测试，系统应用将对比预测结果与实际结果。测试环境如表5-1所示：

表5-1测试环境

Table5-1 test environment

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 项目 | 容量和配置需求 | 数量（台/套） |
| 1 | 数据分析处理集群 | Intel 8Core，128G内存，2T硬盘 | 20 |
| 2 | 采集接口服务器 | PCSERVER,8CPU，32GB内存，500G硬盘 | 1 |
| 3 | MySQL数据库服务器 | PCSERVER,8CPU，32GB内存，500G硬盘 | 1 |
| 4 | 分析结果展现web服务器 | PCSERVER,2CPU，32GB内存，500G硬盘 | 2 |

## 5.1系统测试

(1)系统功能性测试

功能性测试采用黑盒测试方法，通过结合系统具体功能点来进行测试，方法如下:以区域人流监控与预警系统的设计文档和用户手册为依据设计测试，并进行功能点的测试；利用边界值分析方法、等价类划分**[8]**等一些常用软件经常使用的测试方法来对系统各项功能进行测试。表5-2列出该系统的33个功能点及对应的模块名称，并在进行测试时全部覆盖所有的33个功能点。

表5-2商业舆情监控系统的功能点以及测试概况

Table5-2 The system function point and profile

|  |  |  |
| --- | --- | --- |
| 系统功能模块名称 | 功能点 | 测试覆盖并通过的功能点 |
| 账号管理与系统日志模块 | 3 | 3 |
| 客流实时监控与区域人群特征统计模块 | 10 | 10 |
| 客流预测与告警级别阈值设置模块 | 8 | 8 |
| 规则设置与预警响应模块 | 7 | 7 |
| 互联网舆情关键字设置模块 | 5 | 5 |

测试结果分析：各个阶段的测试缺陷量如表5-3所示，在单元测试方面缺陷量相对较少，集成测试缺陷最多，系统测试有所下降，但最终在项目验收时测试全部通过。

表5-3功能测试结果表

Table 5-3 List of function test results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 测试类型\缺陷数(个) | 单元测试 | 集成测试 | 系统测试 | 验收测试 |
| 缺陷数 | 8 | 21 | 13 | 0 |

(2)系统性能测试

系统性能测试采用WebLoad进行并发测试**[9]**，用该工具可以自动执行压力测试，通过模拟真实用户的操作，生成压力负载来测试web的性能。分别模仿5个线程数，10个线程数，20个线程数进行测试，进行多次实验并取平均值。

系统性能测试结果分析：在性能测试方面，通过测试5个线程数，10个线程数，20个线程数测试结果如表5-3所示，通过表中的数据明显可以看出随着用户数量的增多，资源的使用量会越来越大。

表5-3系统性能测试结果

Table 5-3 List of performance test results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 线程数量 | 每次时间间隔 | 平均响应时间(秒) | mysql服务器CPU占用率（平均／最大） | | Web服务器CPU占用率  （平均／最大） | |
| 5 | 0 | 0.127 | 9.24 | 19.32 | 8.42 | 15.93 |
| 10 | 0 | 0.452 | 16.62 | 25.84 | 13.63 | 22.62 |
| 20 | 0 | 0.998 | 22.17 | 29.43 | 24.62 | 27.88 |

系统性能测试结果趋势对比图如图5-1所示：

图5-1性能分析数据图

Fig.5-1 Chart of performance test results

通过测试结果可以看出系统整体质量比较优秀，但web前端性能需要优化，可降低压力从而更进一步提高系统的性能。

## 5.2系统的应用

本系统实现了对区域客流监控与预警，使用者可以在系统中设定告警阈值，从而系统将根据级别进行预警响应。这些功能相对于传统的凭经验预测来说大大提高了工作效率**[10]**。该系统把市民出行的行为与城市地理区域位置相结合，以客流监控、区域人口分析为出发点，利用交通卡出入站闸机AFC系统以及电信OIDD信令数据，经过融合、汇总和分析为该系统提供了有力的数据支撑。系统测试以及六个月的实际上线运行表明，该系统是可行的和有效的。图5-2显示了预测值与实际值的比较之后的一天之内平均相对误差率。

图5-2客流预测效果分析图

Fig.5-2 Effect analysis diagram

2015年5月23日人民广场区域单日内每15分钟短期客流预测的平均绝对百分比误差在6.436%。6:30以前及23:00以后认为是在首末班车时间段内，由于在该时间段内的实际客流数较少，所以首末班车时段内的预测结果的意义不是很大，而非首末班车时间段内的客流预测总体预测结果较为准确，从而说明将时间序列模型应用于区域短时客流量的预测是有效的。

# 6结论

该系统已上线为有关部门提供事前预警和事后分析的支持，试运行半年来情况良好，在早晚高峰以及节假日的客流预测中起到了作用。该系统可实时获取特定区域内人员数量，当预测客流数量激增或出现异常时，可做到提前自动告警，从而达到了监控预警的目的。本系统具有以下几个方面的特点：

1）事前预警：以图形界面展示分析结果。

2）事后分析：以图形界面展示分析结果。

3）数据接口需求：以webservice方式提供电信数据、交通卡以及其他相关数据。

目前，通过处理多维度的数据能够有效对城市区域人流做到大客流监控与预警，有利于事前预警与事后分析，且结果比较准确。但从长远的角度看，如何在现有条件下，优化市民出行的交通线路，实现对有限的城市空间资源更合理地使用，是本系统要进一步深入研究探索的问题。

# 参考文献

[1] 叶霞飞，明瑞利，李忍相. 东京、首尔轨道交通客流成长规律与特征分析[J]. 城市交通，2008(6):16-20.

[2] 韩琳,数据服务平台的服务发布和数据访问研究,华中科技大学, 2012

[3] 高慧.基于混合智能计算的城市交通流预测研究.济南大学, 2008

[4] 锡明，陈必壮，王祥．基于轨道交通网络的大城市综合交通规划理念[J]．城市交通，2010，8(4)：52．

[5]森ビル都市再生プロジェクトチーム. 都市のチカラ：超高層化が生活を豊かにする[M]. 東京都：幻冬舎株式会社，2003.

[6] 张小篆，基于Java与Web模式工作流管理系统[D]，中山大学，2005.

[7] 李辉，刘文军，基于Web的政府公共服务系统开发平台的设计[J]，计算机应用与软件，2008(12),57-62.

[8] 东京都总务局统计部. 东京都统计调查[EB/OL].2005.http://www.toukei.metro.tokyo.jp.

[9] Roy Thomas Fielding. Architectural Styles and the Design of Network-based Software Architectures. UNIVERSITY OF CALIFORNIA, IRVINE [J],2000

[10] STEPHEN C. Traffic prediction using multivariate nonparametric regression［J］. Journal of Transportation Engineering，2003，129(2)：161-168.

[11] K. Morton, M. Balazinska, and D. Grossman. Paratimer: a progress indicator for mapreduce dags[C]. In Proceedings of the 2010 international conference on Management of Data (SIGMOD’10), New York, NY, USA, 2010: 507-518.

1. **作者简介：**倪嘉琦（1991.1-）男，工程硕士，研究方向：软件工程。 [↑](#footnote-ref-1)