基于WIFI探针的商业大数据分析技术

团队：无微不至

成员：周璇 吴小娟 韩霖

日期：2017-08-20

目 录

[摘 要 4](#_Toc12627)

[1概述 5](#_Toc22870)

[2 系统设计 6](#_Toc6073)

[2.1 基于B/S的架构 6](#_Toc28929)

[2.2 数据库设计 6](#_Toc16580)

[2.2.1 sumTable 7](#_Toc13273)

[2.2.2 Table1 8](#_Toc5106)

[2.2.3 Table2 9](#_Toc26189)

[2.2.4 Table3 10](#_Toc31539)

[2.3 Wifi与服务器对接流程 11](#_Toc11468)

[2.4 前台请求数据流程 11](#_Toc19092)

[2.5 后台发送json格式 12](#_Toc3893)

[2.6 前台展示 13](#_Toc1619)

[2.6.1 入店量，客流量 13](#_Toc13611)

[2.6.2 驻店时长，来访周期，客户活跃度 13](#_Toc635)

[2.6.3详细页面客流量 14](#_Toc2485)

[2.6.4 客流量趋势 14](#_Toc28794)

[2.6.5 驻店时长，来访周期，客户活跃度 15](#_Toc10499)

[2.6.6 详细页面入店量 15](#_Toc21072)

[2.6.7 入店率 16](#_Toc20830)

[2.6.8 详细页面来访周期 16](#_Toc18503)

[2.6.9 新老客户比 17](#_Toc6277)

[2.6.10 客户活跃度 17](#_Toc5457)

[2.6.11 详细页面驻店时长 18](#_Toc28934)

[2.6.12 探针一跳出率 18](#_Toc22407)

[2.7 算法思路 19](#_Toc10072)

[2.7.1客流量 19](#_Toc7509)

[2.7.2 入店量 19](#_Toc31934)

[2.7.3 入店率 19](#_Toc1369)

[2.7.4 来访周期 20](#_Toc30260)

[2.7.5 新老顾客 20](#_Toc17930)

[2.7.6 顾客活跃度 20](#_Toc26628)

[2.7.7 驻店时长 21](#_Toc9105)

[2.7.8 跳出率 21](#_Toc19117)

[2.7.9 深访率 22](#_Toc19842)

[3 测试 22](#_Toc3130)

[3.1 Jmeter性能测试 22](#_Toc11099)

摘 要

探针的原理是利用智能设备商WIFI模块所发出的无线广播信号进行设备的感知，就像是网站上的Cookie，他会记录你的访问行为和轨迹。不同的是，通过手机MAC地址所采集的是你的线下行为轨迹，比如喜欢逛什么地方，一周逛几次。相同的是，他都无法直接记录你的个人信息，例如你是谁，手机号多少。

通过对采集MAC地址数据的分析与统计，可以把握门店的客流情况，精准监控客流质量，实时展示客流转化情况，从而帮助检测营销效果，发现潜在机会和改进措施，为便捷、高效精细化运营提供全方位数据参考。

1概述

主要是实现三个方面的功能：一是通过探针设备采集可监测范围内的手机MAC地址、地理位置、与探针距离、时间等信息；二是探针采集的数据可以定时发送到服务端保存；三是利用大数据技术对数据进行人流量等指标的分析。系统应具备以下功能：

**（1）WIFI探针**

探针设备可以在其他渠道（淘宝）购买，也可以自行开发硬件和固件，探针设备可以进行服务端的相关配置（服务端IP、端口、路径、发送数据时间间隔），能够采集MAC地址、地理信息、与探针大概距离、采集时间等信息（不允许采集个人隐私信息，比如手机号等信息），特别针对ANDROID6.0和IOS10版本后的移动终端设备进行测试能否采集到MAC地址。自行开发探针硬件附加10分，自行开发固件附加5分。

**（2）数据采集**

服务端主要接收探针定时发送的数据，将数据保存到数据分析平台待用，文件系统可以使用HDFS或者其他适合的分布式文件系统。接收数据不能有数据丢失或者数据失真，探针每3秒发送一次数据，数据采集并发量不得低于1000台设备，小于1000扣10分，大于1300附加10分。采集的JSON数据结构样例如下：

**（3）数据分析**

基本能够分析以下9大指标，但不仅限于以下9点，参赛选手可以自由发挥，只要在采集到的数据（可以使除探针设备以外的其他数据）基础上的分析合理都可附加创新分，每项5分。能够实时展示结果附加10分。

客流量：店铺或区域整体客流及趋势

入店量：进入店铺或区域的客流及趋势

入店率：进⼊店铺或区域的客流占全部客流的比例及趋势

来访周期：进⼊店铺或区域的顾客距离上次来店的间隔

新老顾客：一定时间段内首次/两次以上进⼊店铺的顾客

顾客活跃度：按顾客距离上次来访间隔,划分为不同活跃度（高活跃度、中活跃度、低活跃度、沉睡活跃度）

驻店时长：进⼊店铺的顾客在店内的停留时长

跳出率：进⼊店铺后很快离店的顾客及占比(占总体客流)

深访率：进⼊店铺深度访问的顾客及占⽐(占总体客流)（可以根据定位轨迹或者停留时长判定）

以上数据指标支持环比和历史对比，并且可以从小时、日

周、月多维度分析。数据分析中的关于范围的界定支持自定义阈值。

2 系统设计

2.1 基于B/S的架构

前台页面采用Bootstrap和echarts框架。前台向后台请求数据。Bootstrap 是一个用于快速开发 Web 应用程序和网站的前端框架。基于 HTML、CSS、JAVASCRIPT 的。echarts是纯Javascript图表库，基于Canvas，底层依赖ZRender，商业产品通用图表库，提供折线图，柱状图，饼图等一系列通用图标的展示方式。Q

后端采用java web设计，Wifi探针向配置的本地服务器的URL以post的形式发送数据。本地运行Servlet的doPost()的方法接受数据，分析数据。

架构图如下所示:

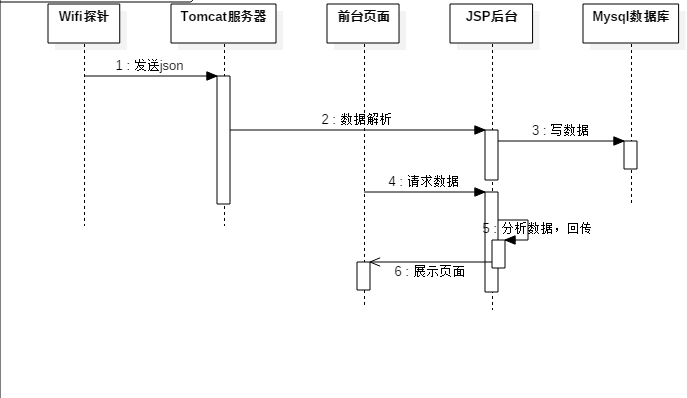


图 1 架构图

2.2 数据库设计

数据分析

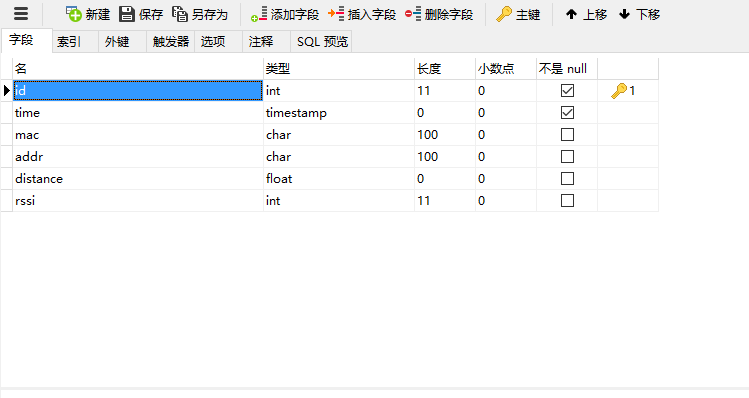
利用JSONObject对json数据进行解析。

数据表设计

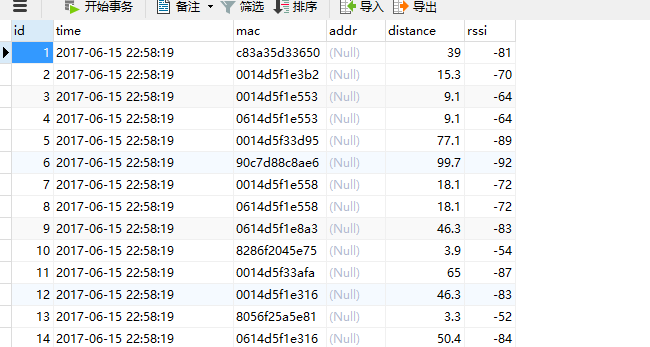
2.2.1 sumTable

存储探针发送的所有json数据，未经加工。

字段为:时间，id，mac，地址，信号强度，距离。

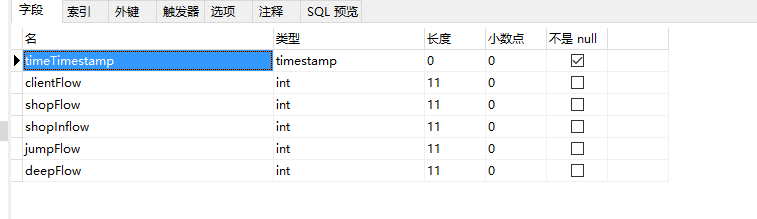


数据展示：

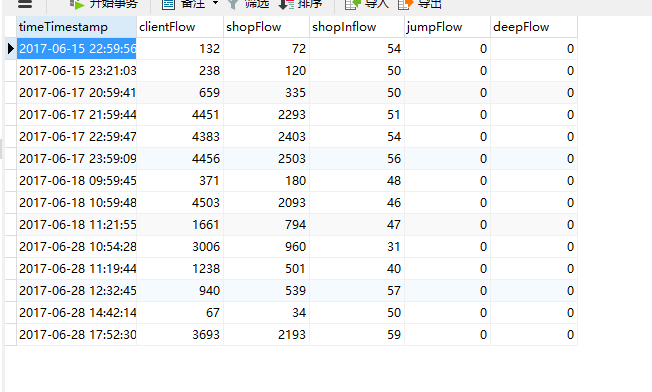


2.2.2 Table1

统计客流量，入店量，入店率，跳出率和深访率。

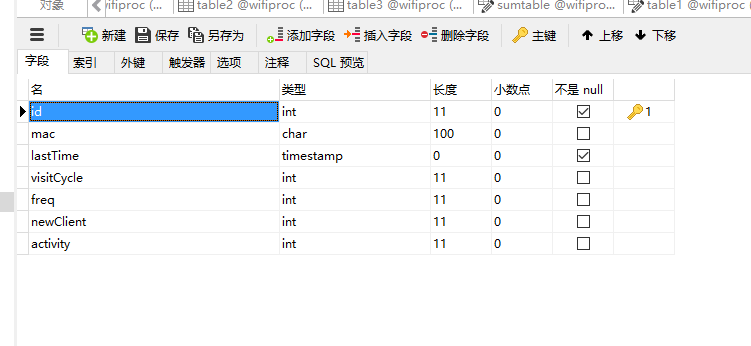


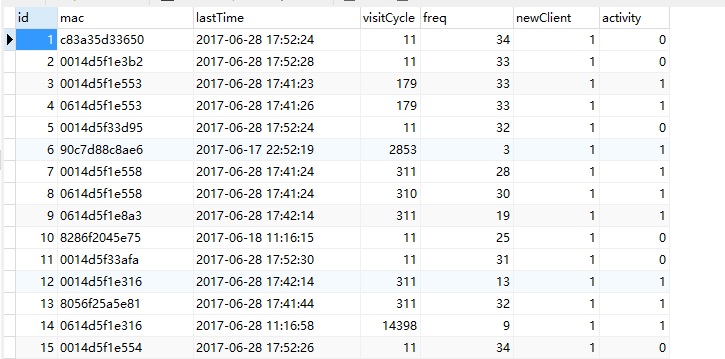
数据展示:



2.2.3 Table2

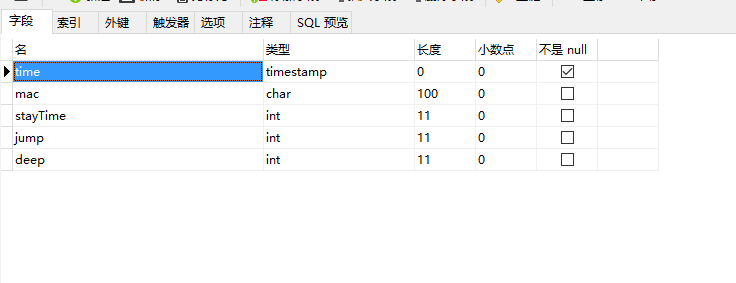
统计驻店时长，来访周期,活跃度，新老客户。

数据展示：

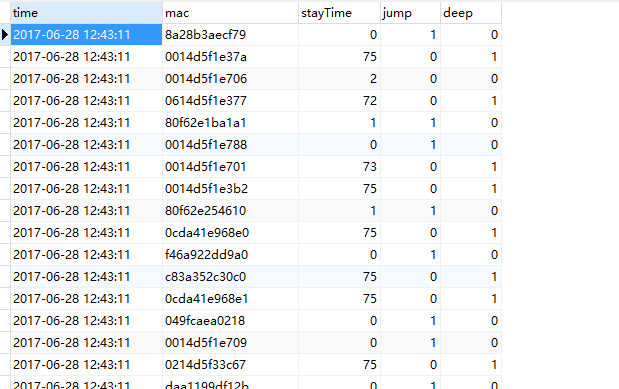


2.2.4 Table3

统计驻店时长，标记是否跳出，是否深访

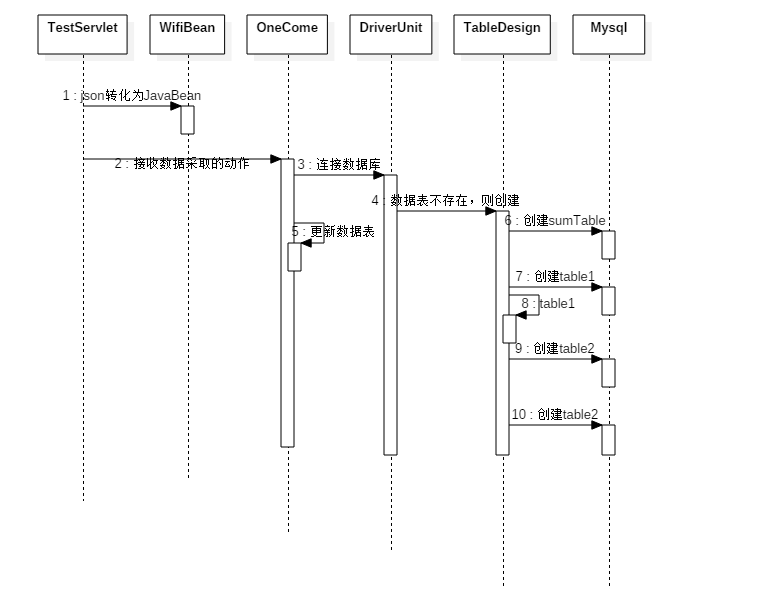


数据展示：



2.3 Wifi与服务器对接流程

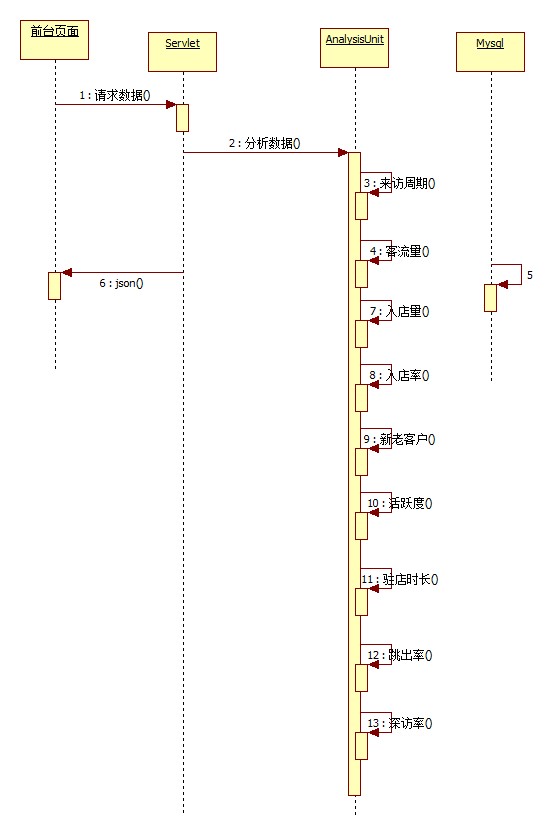
Wifi探针向服务器发送数据，Servlet在其doPost()方法中接收数据，调用OneCome类处理数据，更新数据表。



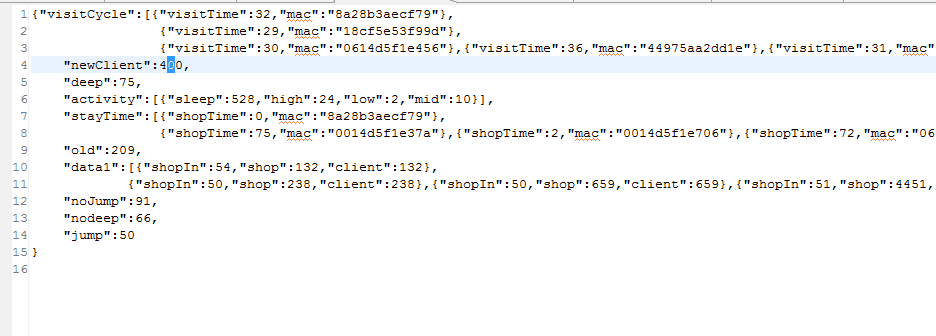
2.4 前台请求数据流程

调用Jquery函数刷新页面，以post的形式想后台servlet请求数据。Servlet调用

AnalysisUnit对数据进行整理，以json的格式传送给前台，前台页面解析后展示。



2.5 后台发送json格式



2.6 前台展示

2.6.1 入店量，客流量



2.6.2 驻店时长，来访周期，客户活跃度



2.6.3详细页面客流量

(客流总量为真实数据，探针一和探针二为模拟数据)



2.6.4 客流量趋势



2.6.5 驻店时长，来访周期，客户活跃度



2.6.6 详细页面入店量

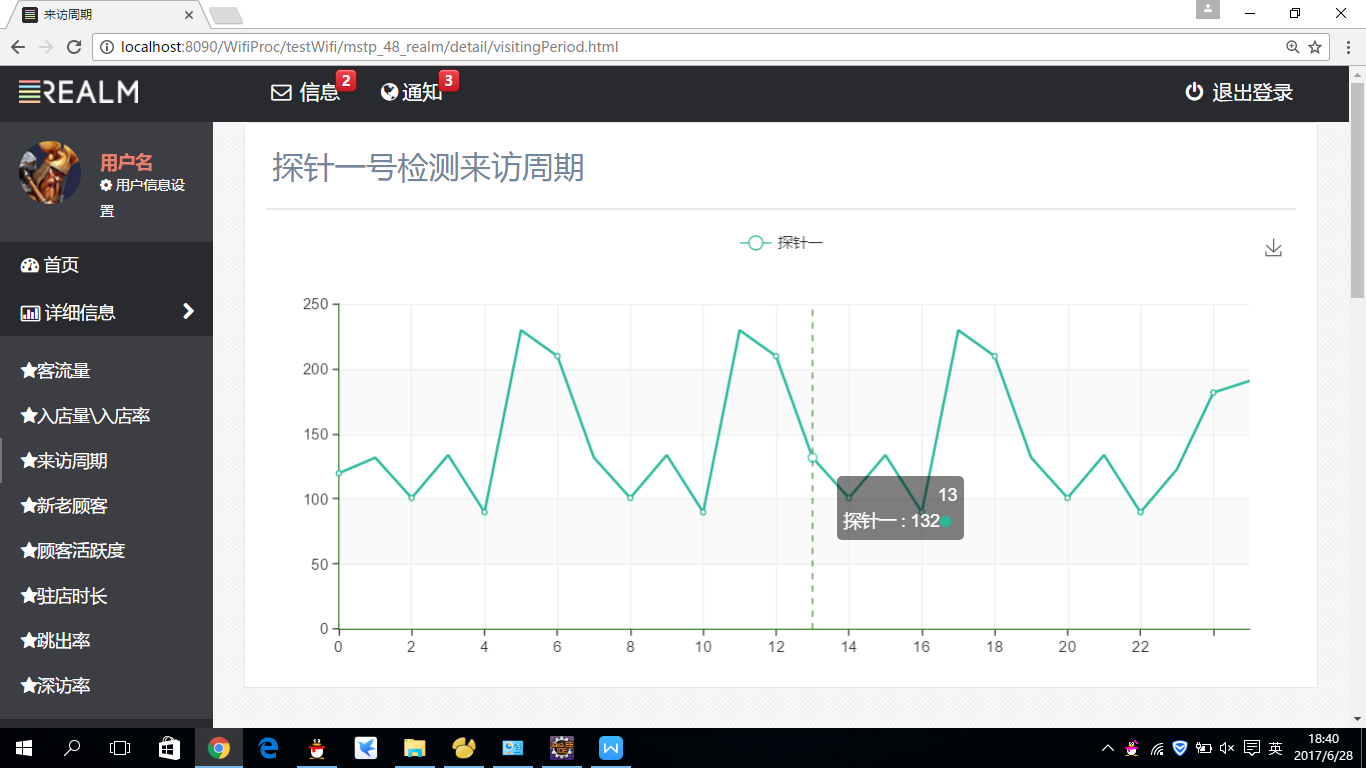
(探针一有效)



2.6.7 入店率

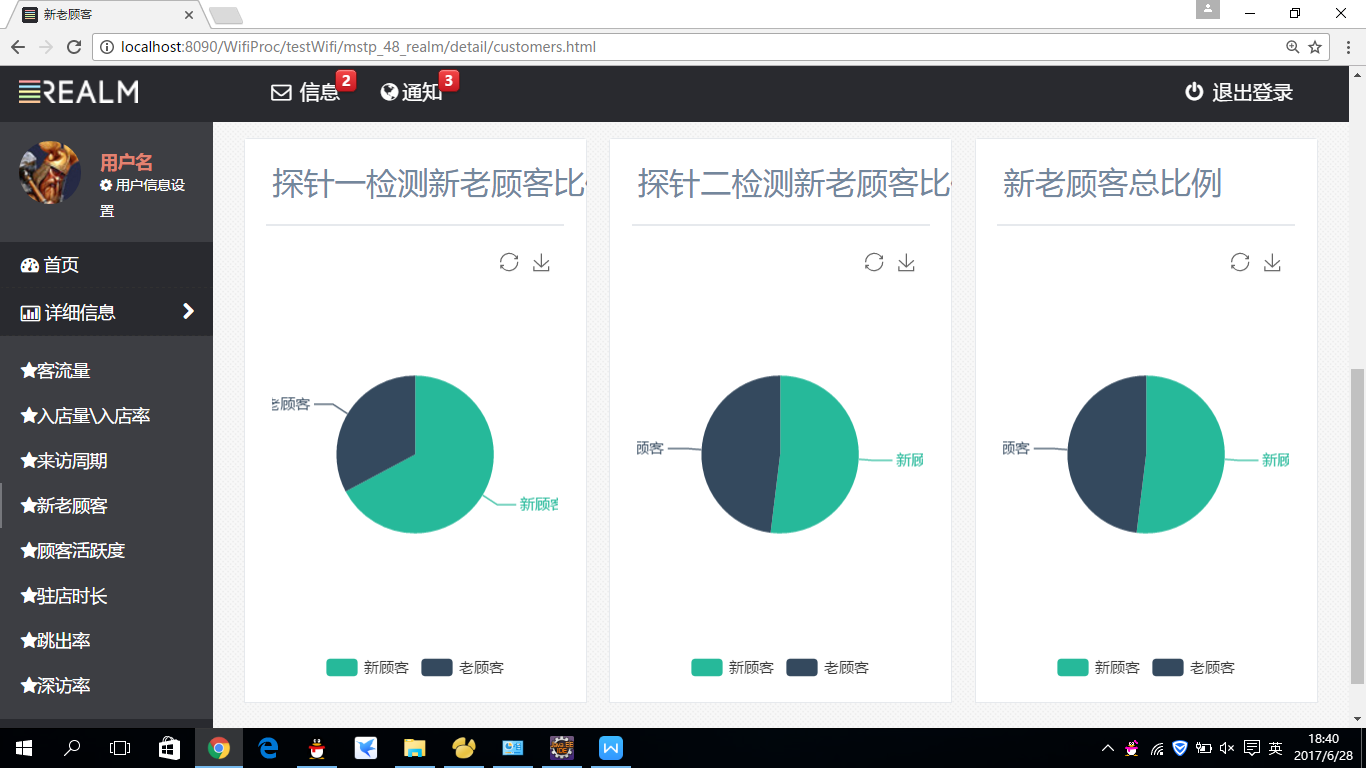


2.6.8 详细页面来访周期



2.6.9 新老客户比

(探针一有效)



2.6.10 客户活跃度

探针一有效)



2.6.11 详细页面驻店时长



2.6.12 探针一跳出率



2.7 算法思路

2.7.1客流量

店铺或区域整体客流及趋势

以小时为单位，统计时间段内不同MAC值的数量。

因为只有一个探针，如果要统计客流量和入店量的不同，只有一个距离作为指标。

设置入店量范围下限，如100m和客流量距离下限，如50m。

筛选MAC:使用数组(PHP中好像没有不允许重复值的数据结构),每当检测到一个MAC将其与数组中的其他值做对比，若没有此MAC，则将其加入数组，客流量加1.若此MAC的距离小于50m，则客流量与入店量均加1;若MAC的距离大于50m并且小于100m，则客流量加1;若大于100m，则此MAC不加入数组，不做任何操作。

2.7.2 入店量

进入店铺或区域的客流及趋势

2.7.3 入店率

进⼊店铺或区域的客流占全部客流的比例及趋势

入店量/客流量

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时间 | 客流量 | 入店量 | 入店率 | 跳出比 |
| 2017-04-14 12：00 | 490 | 234 | 0.33\*100 | 0.44\*100 |
| 2017-04-17 13：00 | 378 | 123 | 0.40\*100 | 0.55\*100 |

流量统计图

2.7.4 来访周期

进⼊店铺或区域的顾客距离上次来店的间隔

计算来访频率:

某段时间内同一MAC地址出现的次数，按降序排序。

SELECT MAC,COUNT(\*) FROM detail BETWEEN 2017-04-15 12：00 AND 2017-04--17 13：00 GROUP BY MAC HAVING COUNT(\*) >1 order by count DESC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MAC | 上次来的时期 | 来访周期 | 来访频率 | 新老客户 | 活跃度 |
|  | 2017-04-16 12:00 | 30 | 23 | 1 | 3 |
|  | 2017-04-17 13:00 | 113 | 1 | 0 | 1 |

MAC活跃度表

每当检测到一个MAC地址，检测该表的’上次来的时期’字段，设置一个间隔值(相同MAC在同一时间内会向wifi探针发送多次信号，因此时间间隔太小的情况下就代表是同一次访问)，若当前MAC来访时间-上次来访时间<间隔值，则不更新此字段，若大于，则本次来访时间记录为上次来访时间。

2.7.5 新老顾客

一定时间段内首次/两次以上进⼊店铺的顾客

根据来访频率来计算是否新老客户，若来访频率>1为老客户，字段属性为1;若来访频率=1

为新客户，字段属性为0.

2.7.6 顾客活跃度

按顾客距离上次来访间隔,划分为不同活跃度（高活跃度、中活跃度、低活跃度、沉睡活跃度）

根据来访周期的时间划分为不同活跃度。

高活跃度:4

中活跃度:3

低活跃度:2

沉睡活跃度:1

2.7.7 驻店时长

进⼊店铺的顾客在店内的停留时长

假设某人在商店里驻留时常不超过一个小时，在此时间段内设置一个开始时间和一个结束时间。比如12:00--13:00。设置一个距离下限，若客户距离在此距离内，则代表用户进店，记录进店时间。检测某个MAC地址在此时间段内的像WIFI探针发送的每个信号的距离，设置一个距离上限，若超过此上限，则代表用户离店，记录离店时间，驻留时长=进店时间-离店时间。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时间 | MAC | 驻留时长 | 是否跳出 | 是否深访 |
| 2017-04-17 12：00 |  |  | 0 | 1 |
| 2017-04-17 13：00 |  |  | 1 | 0 |

驻店时间统计表

2.7.8 跳出率

进⼊店铺后很快离店的顾客及占比(占总体客流)

同上，设置一个驻留时间下限，若驻留时间小于该下限，则是属于进入店铺后很快离店的客户。总体客流是指同时间段内客流量。每当检测完一个时间段就将新的跳出率插入到表中。计算跳出率：检测同一时段里’是否跳出’字段，若为1则为跳出，为0则不跳出，总的跳出占总的MAC的比值。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 时间 | 客流量 | 入店量 | 入店率 | 跳出比 | 深访率 |
| 2017-04-14 12：00 | 490 | 234 | 0.33\*100 | 0.44\*100 | 0.22\*100 |
| 2017-04-17 13：00 | 378 | 123 | 0.40\*100 | 0.55\*100 | 0.31\*100 |

跳出比统计表

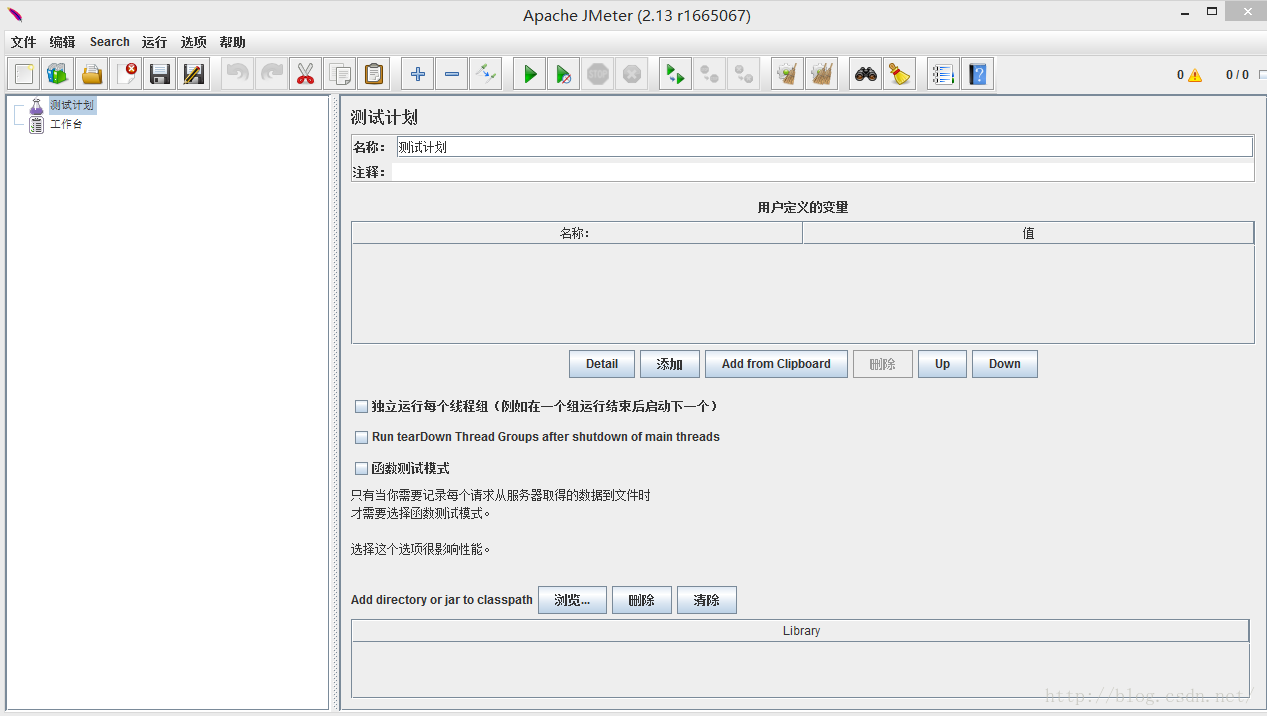
2.7.9 深访率

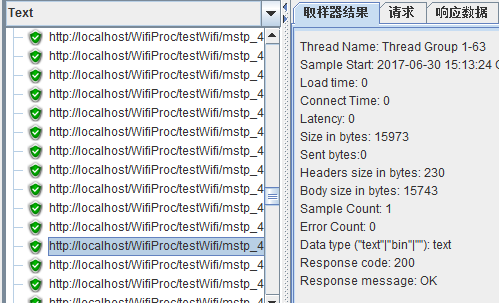
进⼊店铺深度访问的顾客及占⽐(占总体客流)（可以根据定位轨迹或者停留时长判定）

设置一个深度访问的时间限制.若超过此限制，客户为深度访问客户,’是否深访’字段设置为1；若低于，不是深度访问客户，属性值设置为0.

3 测试

3.1 Jmeter性能测试





数据请求的jmeter中结果树的监听结果如图

