

# 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.....	3
1 概述.....	4
2 系统设计.....	5
2.1 基于 B/S 的架构.....	5
2.2 数据库设计.....	5
2.2.1 sumTable.....	5
2.2.2 Table1.....	7
2.2.3 Table2.....	8
2.2.4 Table3.....	9
2.3 Wifi 与服务器对接流程.....	10
2.4 前台请求数据流程.....	10
2.5 后台发送 json 格式.....	12
2.6 前台展示.....	12
2.6.1 入店量, 客流量.....	12
2.6.2 驻店时长, 来访周期, 客户活跃度.....	12
2.6.3 详细页面客流量.....	13
2.6.4 客流量趋势.....	13
2.6.5 驻店时长, 来访周期, 客户活跃度.....	14
2.6.6 详细页面入店量.....	14
2.6.7 入店率.....	15
2.6.8 详细页面来访周期.....	15
2.6.9 新老客户比.....	16
2.6.10 客户活跃度.....	16
2.6.11 详细页面驻店时长.....	17
2.6.12 探针一跳出率.....	17
2.7 算法思路.....	18
2.7.1 客流量.....	18
2.7.2 入店量.....	18
2.7.3 入店率.....	18
2.7.4 来访周期.....	19
2.7.5 新老顾客.....	19
2.7.6 顾客活跃度.....	20
2.7.7 驻店时长.....	20
2.7.8 跳出率.....	20
2.7.9 深访率.....	21
3 测试.....	21
3.1 Jmeter 性能测试.....	21

## 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：

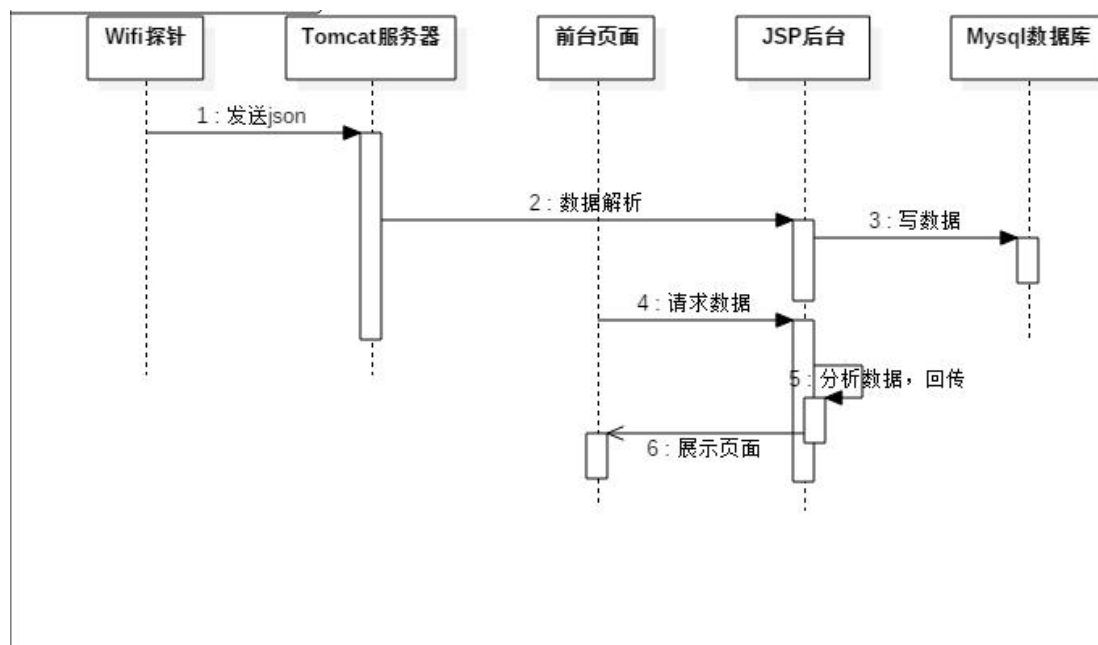


图 1 架构图

### 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。

字段	索引	外键	触发器	选项	注释	SQL 预览
名	类型	长度	小数点	不是 null		
id	int	11	0	<input checked="" type="checkbox"/>	1	
time	timestamp	0	0	<input checked="" type="checkbox"/>		
mac	char	100	0	<input type="checkbox"/>		
addr	char	100	0	<input type="checkbox"/>		
distance	float	0	0	<input type="checkbox"/>		
rss	int	11	0	<input type="checkbox"/>		

数据展示：

id	time	mac	addr	distance	rss
1	2017-06-15 22:58:19	c83a35d33650	(Null)	39	-81
2	2017-06-15 22:58:19	0014d5f1e3b2	(Null)	15.3	-70
3	2017-06-15 22:58:19	0014d5f1e553	(Null)	9.1	-64
4	2017-06-15 22:58:19	0614d5f1e553	(Null)	9.1	-64
5	2017-06-15 22:58:19	0014d5f33d95	(Null)	77.1	-89
6	2017-06-15 22:58:19	90c7d88c8ae6	(Null)	99.7	-92
7	2017-06-15 22:58:19	0014d5f1e558	(Null)	18.1	-72
8	2017-06-15 22:58:19	0614d5f1e558	(Null)	18.1	-72
9	2017-06-15 22:58:19	0614d5f1e8a3	(Null)	46.3	-83
10	2017-06-15 22:58:19	8286f2045e75	(Null)	3.9	-54
11	2017-06-15 22:58:19	0014d5f33afa	(Null)	65	-87
12	2017-06-15 22:58:19	0014d5f1e316	(Null)	46.3	-83
13	2017-06-15 22:58:19	8056f25a5e81	(Null)	3.3	-52
14	2017-06-15 22:58:19	0614d5f1e316	(Null)	50.4	-84

2.2.2 Table1

统计客流量，入店量，入店率，跳出率和深访率。

字段	索引	外键	触发器	选项	注释	SQL 预览				
名					类型	长度	小数点	不是 null		
timeTimestamp					timestamp	0	0	<input checked="" type="checkbox"/>		
clientFlow					int	11	0	<input type="checkbox"/>		
shopFlow					int	11	0	<input type="checkbox"/>		
shopInflow					int	11	0	<input type="checkbox"/>		
jumpFlow					int	11	0	<input type="checkbox"/>		
deepFlow					int	11	0	<input type="checkbox"/>		

数据展示:

timeTimestamp	clientFlow	shopFlow	shopInflow	jumpFlow	deepFlow
2017-06-15 22:59:56	132	72	54	0	0
2017-06-15 23:21:03	238	120	50	0	0
2017-06-17 20:59:41	659	335	50	0	0
2017-06-17 21:59:44	4451	2293	51	0	0
2017-06-17 22:59:47	4383	2403	54	0	0
2017-06-17 23:59:09	4456	2503	56	0	0
2017-06-18 09:59:45	371	180	48	0	0
2017-06-18 10:59:48	4503	2093	46	0	0
2017-06-18 11:21:55	1661	794	47	0	0
2017-06-28 10:54:28	3006	960	31	0	0
2017-06-28 11:19:44	1238	501	40	0	0
2017-06-28 12:32:45	940	539	57	0	0
2017-06-28 14:42:14	67	34	50	0	0
2017-06-28 17:52:30	3693	2193	59	0	0

### 2.2.3 Table2

统计驻店时长，来访周期,活跃度，新老客户。

对象   < wifiproc (...)   table2 @wifiproc (...)   table3 @wifiproc (...)   sumtable @wifipro...   table1 @wifiproc (...)						
新建 保存 另存为 添加字段 插入字段 删除字段 主键 上移 下移						
字段 索引 外键 触发器 选项 注释 SQL 预览						
名	类型	长度	小数点	不是 null		
id	int	11	0	<input checked="" type="checkbox"/>	1	
mac	char	100	0	<input type="checkbox"/>		
lastTime	timestamp	0	0	<input checked="" type="checkbox"/>		
visitCycle	int	11	0	<input type="checkbox"/>		
freq	int	11	0	<input type="checkbox"/>		
newClient	int	11	0	<input type="checkbox"/>		
activity	int	11	0	<input type="checkbox"/>		

数据展示:

id	mac	lastTime	visitCycle	freq	newClient	activity
1	c83a35d33650	2017-06-28 17:52:24	11	34	1	0
2	0014d5f1e3b2	2017-06-28 17:52:28	11	33	1	0
3	0014d5f1e553	2017-06-28 17:41:23	179	33	1	1
4	0614d5f1e553	2017-06-28 17:41:26	179	33	1	1
5	0014d5f33d95	2017-06-28 17:52:24	11	32	1	0
6	90c7d88c8ae6	2017-06-17 22:52:19	2853	3	1	1
7	0014d5f1e558	2017-06-28 17:41:24	311	28	1	1
8	0614d5f1e558	2017-06-28 17:41:24	310	30	1	1
9	0614d5f1e8a3	2017-06-28 17:42:14	311	19	1	1
10	8286f2045e75	2017-06-18 11:16:15	11	25	1	0
11	0014d5f33afa	2017-06-28 17:52:30	11	31	1	0
12	0014d5f1e316	2017-06-28 17:42:14	311	13	1	1
13	8056f25a5e81	2017-06-28 17:41:44	311	32	1	1
14	0614d5f1e316	2017-06-28 11:16:58	14398	9	1	1
15	0014d5f1e554	2017-06-28 17:52:26	11	34	1	0



### 2.2.4 Table3

统计驻店时长，标记是否跳出，是否深访

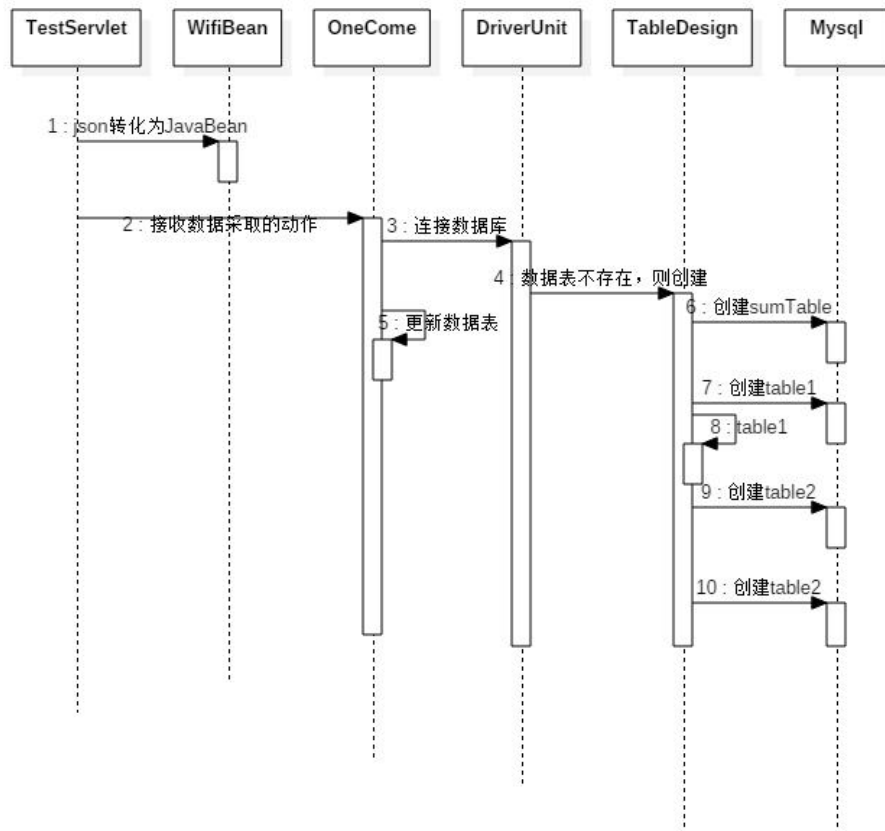
字段							索引	外键	触发器	选项	注释	SQL 预览
名		类型		长度	小数点	不是 null						
▶	time	timestamp		0	0	<input checked="" type="checkbox"/>						
	mac	char		100	0	<input type="checkbox"/>						
	stayTime	int		11	0	<input type="checkbox"/>						
	jump	int		11	0	<input type="checkbox"/>						
	deep	int		11	0	<input type="checkbox"/>						

数据展示：

time	mac	stayTime	jump	deep
2017-06-28 12:43:11	8a28b3aecf79	0	1	0
2017-06-28 12:43:11	0014d5f1e37a	75	0	1
2017-06-28 12:43:11	0014d5f1e706	2	0	0
2017-06-28 12:43:11	0614d5f1e377	72	0	1
2017-06-28 12:43:11	80f62e1ba1a1	1	1	0
2017-06-28 12:43:11	0014d5f1e788	0	1	0
2017-06-28 12:43:11	0014d5f1e701	73	0	1
2017-06-28 12:43:11	0014d5f1e3b2	75	0	1
2017-06-28 12:43:11	80f62e254610	1	1	0
2017-06-28 12:43:11	0cda41e968e0	75	0	1
2017-06-28 12:43:11	f46a922dd9a0	0	1	0
2017-06-28 12:43:11	c83a352c30c0	75	0	1
2017-06-28 12:43:11	0cda41e968e1	75	0	1
2017-06-28 12:43:11	049fcaea0218	0	1	0
2017-06-28 12:43:11	0014d5f1e709	0	1	0
2017-06-28 12:43:11	0214d5f33c67	75	0	1
2017-06-28 12:43:11	daa1199df12h	0	1	0

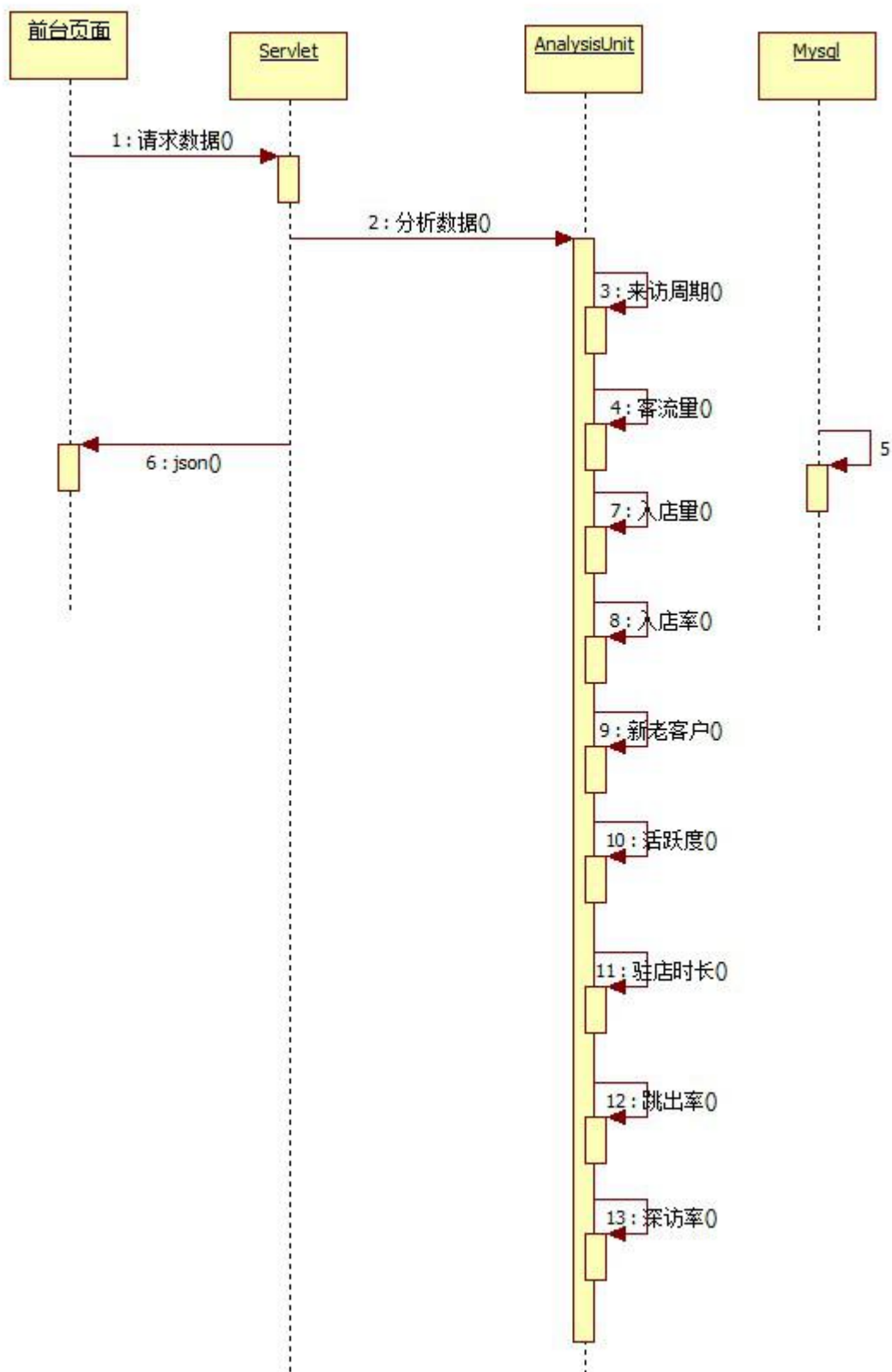
### 2.3 Wifi 与服务器对接流程

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

```
1 {"visitCycle":[{"visitTime":32,"mac":"8a28b3aecf79"},
2               {"visitTime":29,"mac":"18cf5e53f99d"},
3               {"visitTime":30,"mac":"0614d5f1e456"}, {"visitTime":36,"mac":"44975aa2dd1e"}, {"visitTime":31,"mac"
4 "newClient":40,
5 "deep":75,
6 "activity":[{"sleep":528,"high":24,"low":2,"mid":10}],
7 "stayTime":[{"shopTime":0,"mac":"8a28b3aecf79"},
8             {"shopTime":75,"mac":"0014d5f1e37a"}, {"shopTime":2,"mac":"0014d5f1e706"}, {"shopTime":72,"mac":"06
9 "old":209,
10 "data1":[{"shopIn":54,"shop":132,"client":132},
11          {"shopIn":50,"shop":238,"client":238}, {"shopIn":50,"shop":659,"client":659}, {"shopIn":51,"shop":4451,
12 "noJump":91,
13 "nodeep":66,
14 "jump":50
15 }
16 }
```

## 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

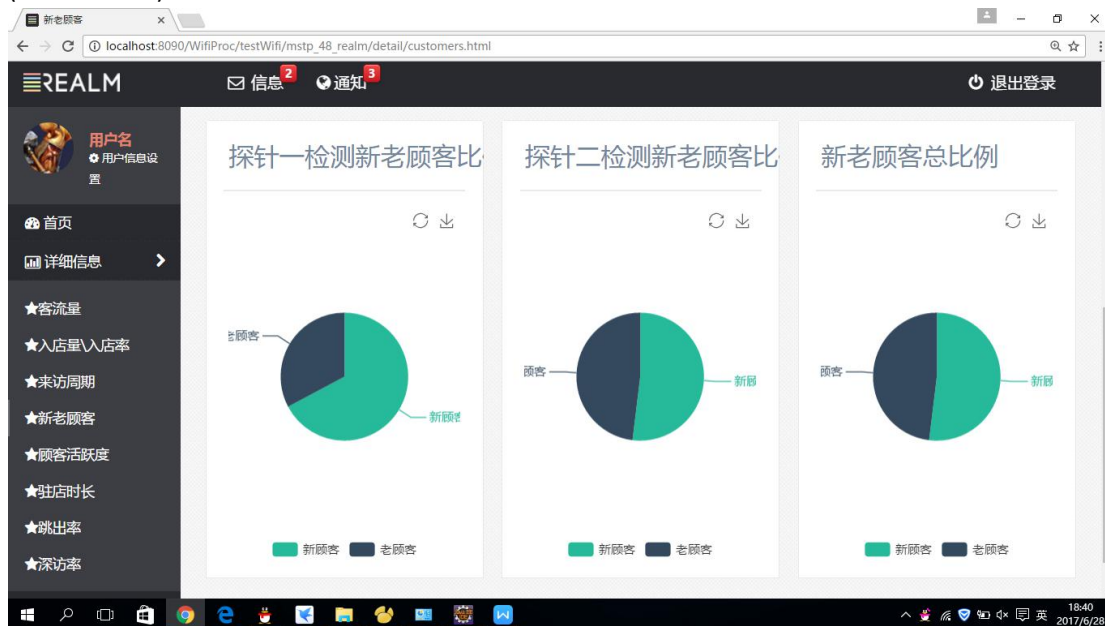


## 2.6.8 详细页面来访周期





## 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进店量的不同，只有一个距离作为指标。

设置进店量范围下限，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，则客流量与进店量均加 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 进店量

进入店铺或区域的客流及趋势

### 2.7.3 进店率

进入店铺或区域的客流占全部客流的比例及趋势

进店量/客流量

时间	客流量	进店量	进店率	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问隔,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:	490	234	0.33*100	0.44*	0.22*

00				100	100
2017-04-17 13:00	378	123	0.40*100	0.55*	0.31*
				100	100

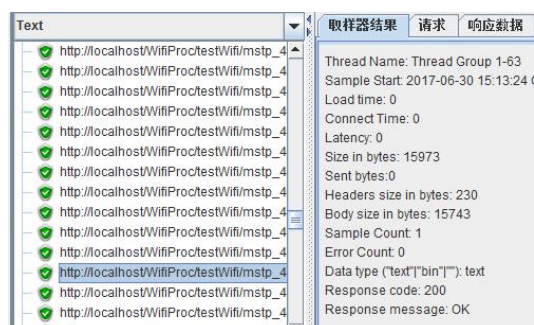
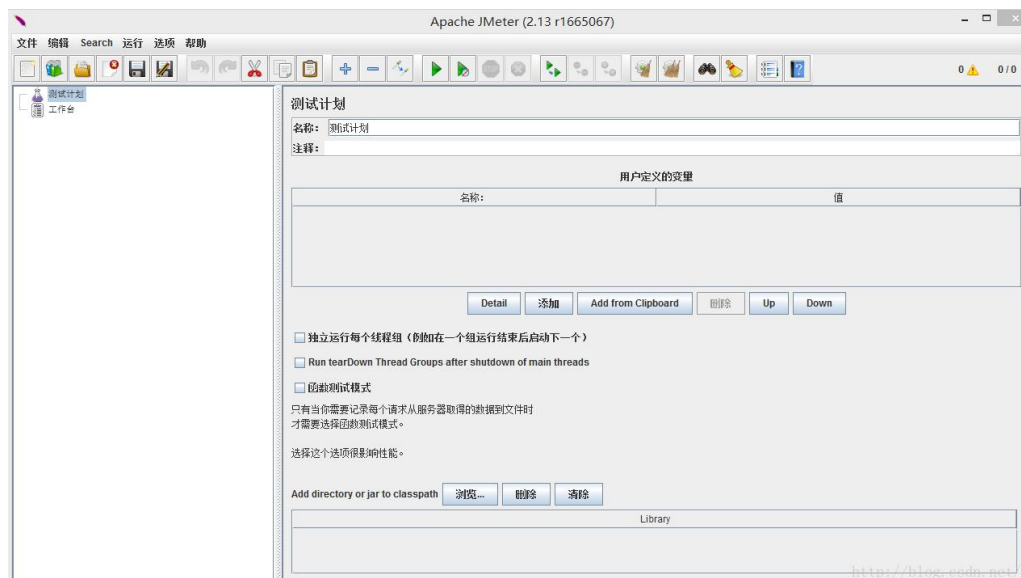
跳出比统计表

### 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

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Max	Error %	Throughput	Received ...	Sent KB/sec
http://local...	100	0	1	1	1	1	0	6	0.00%	33.5/sec	317.48	0.00
http://local...	100	0	1	1	1	3	0	4	0.00%	33.6/sec	523.81	0.00
总体	200	0	1	1	1	3	0	6	0.00%	66.9/sec	839.54	0.00

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Max	Error %	Throughput	Received ...	Sent KB/sec
http://local...	2000	0	1	1	1	5	0	82	0.00%	377.0/sec	5880.83	0.00
http://local...	1000	0	0	1	1	10	0	50	0.00%	188.6/sec	1799.90	0.00
http://local...	1000	0	0	1	1	4	0	26	0.00%	188.7/sec	1802.05	0.00
总体	4000	0	0	1	1	5	0	82	0.00%	754.0/sec	9480.06	0.00