人流预测方案

****一、目前预测方案****

**目前状况由于数据量比较少而且不全，所以目前暂且认为流量数据没有周期性问题，数据按每天各个时间点为单位，预测未来各个时间点的数据，分为按小时和按分钟为单位进行预测。**

**预测算法为迭代加权，这样能考虑到历史的每一个数据并赋予权重迭代预测到最近一天的数据。这种算法在数据量少的情况下能达到比较好的预测效果,具体计算公式如下:**



：时间点n预测值

:时间点n-1 实际值

:时间点n-1预测值

h:实际值权重

l:历史数据权重

****二、后期预测方案****

**1、影响人流量因素**

1. **天气因素**

天气方面关注两个方面 温度和是否下雨

需要获取历史数据观察表如下：

|  |  |  |
| --- | --- | --- |
| **室外温度** | **是否下雨** | **人流量** |
|  |  |  |
|  |  |  |
|  |  |  |

**天气数据来源：**

**以深圳为例子：**

**请求连接：**<http://www.weather.com.cn/data/sk/101280601.html>

**返回结果jason 内容如下：**

{"weatherinfo":{"city":"深圳","cityid":"101280601","temp":"21","WD":"南风","WS":"1级","SD":"31%","WSE":"1","time":"17:05","isRadar":"1","Radar":"JC\_RADAR\_AZ9755\_JB","njd":"暂无实况","qy":"1014","rain":"0"}}

**量化数据：**

**温度:** **摄氏度**

**是否下雨：无、小雨、中雨、大雨量化为**

|  |  |  |  |
| --- | --- | --- | --- |
| **无** | **小雨** | **中雨** | **大雨** |
| **0** | **0.1** | **0.5** | **1** |

**单因素分析:**

**1）天气与人流量关系 忽略下雨因素，观察数据符合哪种关系，做分段线性回归分析**

**2）是否下雨与人流量关系 ，观察数据符合哪种关系，做回归分析**

**多因素分析:**

**引入两个因素(温度和是否下雨)同时观察，做逻辑回归分析。**

1. **节假日**

**由于不同节假日对人流量影响不同，但是总体上都应该会导致人流有所上升趋势，特别是在峰值部分波动幅度应该是大于其他时间点的波动幅度，预设图大概如下：**

1. **获取日期是否为节假日**

**请求API：**<http://www.easybots.cn/api/holiday.php?d=20161225>

**d = 日期 返回json 结果** {"20161225":"1"}

**加入万年历API(目前市场上api 都需要钱)**

**按不同节假日对人流量的影响进行波动值预设，前期可以设置大致波动值可跟海关了解，后期观察数据做分析得出各节假日波动值大小，计算公式如下：**



:假节日时间点n预测值

:普通日期时间点n预测值

:节假日波动值

波动值表如下：

|  |  |
| --- | --- |
| **节假日** | **波动值** |
| **春节** |  |
| **端午节** |  |
| **国庆节** |  |
| **圣诞节** |  |
| **…** |  |

**2、预测频率**

目前状况：目前预测频率为一天一次，会导致忽略了特殊情况因素，比如上一小时突发事件产生，会导致下一小时(或未来几小时)人流量的上升或下跌，这样就会导致当天后面预测数据与真实数据的偏差。

解决方法：先做前期预测即每天都会产生一次粗略预测效果，后加入修正程序，为每小时对预测结果进行修正，取当天时间点前1小时变化趋势作为考量依据。

具体计算公式如下：



U(n-1):上一小时预测值浮动百分比

P(n-1):上一小时预测值

S(n-1):上一小时实际



PR(n) ：修正后预测值

**P(n): 粗略预测值**

**因为事件的发生对人流量的影响会衰减，即当前小时受前一小时事件的影响要小于事件对前一个小时的影响程度，所以波动程度也应该小于前一个小时，也可认为前一个小时的波动程度有置信度，我们暂定0.9,所以公式变为：**



3、数据选取

预测数据的选取，由于比较长远的数据对近期的参考意义会比较少，甚至会影响到近期的预测效果，所以数据选取区间选为最近3个月，同一维度数据大于三组即可引入周期性，以一个星期为变化周期，对历史数据进行周期选取。**数据会剔除异常值（包括特殊情况人流、节假日人流、程序异常值），因为这些数据会对普通日期预测效果干预。**