### 4.2.1数据流定义

数据流包括开始从WIFI探针上读取的原始数据，和进行一步步分析处理得到各类有效数据流，主要包括原始数据流、结构化数据流、模型数据流、历史数据流、可视化数据流（分为历史数据流与实时数据流、管理层与分店数据流）。

|  |
| --- |
| 数据流名称：原始数据 |
| 描述：探针采集到的数据  组成：mac地址+时间戳+信号强度+手机距离嗅探器的距离+经纬度+地址  来源：探针  终点：采集数据 |

|  |
| --- |
| 数据流名称：结构化数据 |
| 描述：系统分类统计的数据  组成：结构化 mac 地址+结构化时间戳+结构化手机距离+经纬度  来源：采集数据  终点：存储至HBase |

|  |
| --- |
| 数据流名称：模型数据 |
| 描述：选用参考模型的参数  组成：历史模型公式+当前客流量+当前时间+预计客流量  来源：训练模型  终点：存储至HBase |

|  |
| --- |
| 数据流名称：历史数据 |
| 描述：之前采集到的数据  组成：MAC地址+时间戳+手机距离+地址  来源：HBase  终点：训练模型 |

|  |
| --- |
| 数据流名称：可视化数据 |
| 描述：将分店面的数据以表报的格式呈现  组成：客流量+入店量+活跃度等分时段对比显示  来源：数据可视化  终点：可视化数据（送至WEB显示） |

### 4.2.2数据存储定义

数据存储主要是对有价值数据进行存储以便后续使用，我们采用HBase分布式数据库（NOSQL），主要包括原始数据存储、数据备份（以便训练模型及查看历史）、模型数据存储、可视化数据存储。

|  |
| --- |
| 数据名：原始数据 |
| 描述：存储各个分店各次实时采集到的数据  组成：mac 地址+时间戳+信号强度+地理位置  相关处理：采集数据 分析访客活跃度 数据可视化 |

|  |
| --- |
| 数据名：数据备份 |
| 描述：存储所有探针采集到的数据，包括各个分店的结构化历史数据  组成：店名+顾客MAC+时间戳+手机距离+地理位置  相关处理：采集数据 训练模型 |

|  |
| --- |
| 数据名：模型数据 |
| 描述：存储模型参数和模型调优的参数  组成：历史模型公式+当前客流量+当前时间+模型公式  相关处理：分析访客活跃度、客流量等 数据可视化 训练模型 |

|  |
| --- |
| 数据名：可视化数据 |
| 描述：存储各个分店各次实时分析得到的入店量、活跃度、访问周期等数据  组成：店名+时间戳+入店量 / 店名+时间戳+活跃度 …….  相关处理：数据可视化 |

### 4.2.3处理逻辑定义

处理逻辑主要是对数据流的操作，包括采集数据逻辑、分析入店量逻辑、预测趋势逻辑、训练模型逻辑。

|  |
| --- |
| 处理名：采集数据 |
| 描述：当顾客进入商场时，采集顾客手机的 MAC 地址、进店时间、信号强度等数据  输入数据流：原始数据  输出数据流：结构化数据  存取的数据库：原始数据、数据备份  处理逻辑：While{探针检测到原始数据}do  存入数据备份  原始数据表中添加对应记录 |

|  |
| --- |
| 处理名：分析入店量 |
| 描述：对采集到的结构化数据进行整理分析，通过顾客距离判断是否入店  输入数据流：结构化数据  输出数据流：各个店的结构化数据  存取的数据库：结构化数据、各个店的入店记录  处理逻辑：if（到某一探针的距离小于阈值且到另外一个探针的距离小于阈值）  if(该用户不在这个店的入店记录)  结构化数据添加到对应记录  else  if(该用户在这个店的入店记录)  删除对应记录并添加该顾客的访问时间 |

|  |
| --- |
| 处理名：预测趋势 |
| 描述：结合当前入店量及历史入店量，利用模型预测未来访店量  输入数据流：采集的数据和模型数据  输出数据流：入店量趋势预测的数据  存取的数据库：原始数据、模型数据、可视化数据  处理逻辑：while（采集到数据）do  对比历史入店量  用模型计算未来访问量 |

|  |
| --- |
| 处理名：训练模型 |
| 描述：根据历史数据，调整分析预测模型  输入数据流：历史数据  输出数据流：模型数据  存取的数据库：数据备份、模型数据  处理逻辑：if（模型预测结果 ！=当前入店量）  调整训练模型 |

### 4.2.4外部实体定义

|  |
| --- |
| 外部实体名：探针 |
| 描述：收集顾客的相关数据包括MAC 地址，时间戳，信号强度等  有关数据流：原始数据 |

|  |
| --- |
| 外部实体名：管理层 |
| 描述：通过系统对各个店铺的客流量、入店量、活跃度等的预测和当前情况，进行对营销方案的制定和修改工作  有关数据流：可视化数据 |

|  |
| --- |
| 外部实体名：分店 |
| 描述：接收本店的顾客实时入店量、活跃度等，及可视化历史数据（包括来访时间、来访时长及各种关于顾客的简单统计信息）并一直显示总体可视化数据（包括今日客流量、跳出率等）  有关数据流：进店顾客可视化数据 |