# 一．数据源

电子围栏表 EFENCE*DETECT*INFO需要的字段，此数据由dts从snowball(mppdb)同步到hive， 算法可直接从hive读取

|  |  |  |
| --- | --- | --- |
| 字段名称 | 字段描述 | 备注 |
| MSISDN | 手机号码 |  |
| IMSI | 国际移动用户识别码 |  |
| IMEI | 国际移动设备标识 |  |
| MAC | MAC |  |
| DEVICE\_ID | 设备编号 |  |
| JGSK | 经过日期时间 | yyMMddHHmmss格式化 |

# 二．需要计算的关系

|  |  |
| --- | --- |
| 关系名称 | 关系代码 |
| 共用手机号的IMEI | 08-0003-0001-02 |
| 共用手机号的MAC | 08-0003-0002-02 |
| 共用IMEI的手机号 | 08-0003-0003-02 |
| 共用IMEI的IMSI | 08-0003-0004-02 |
| 共用MAC的手机号 | 08-0003-0005-02 |
| 共用MAC的IMSI | 08-0003-0006-02 |
| 共用IMSI的IMEI | 08-0003-0007-02 |
| 共用IMSI的MAC | 08-0003-0008-02 |

# 算法

1. 帧码剔除异常值（如000,0000,00000000等），并对每个帧码（如imsi与imei的绑定关系）的每天使用情况，包括使用次数，最大最小发生时间等进行聚合，。
2. 通过共用设备进行关联，计算共现的概率相似度。相似度（similarity）公式如下：

IMG_256

timediff：为时间间隔，count为次数

w1,w2为时间维度和次数维度的权重，这里取w1=0.5,w2=0.5 可以根据实际情况进行调节,这里=0.0127,IMG_256=0.05

# 数据情况

九月1号到11月5号的电围表数据情况（单位：个）

|  |  |  |  |
| --- | --- | --- | --- |
| msisdn | mac | imsi | imei |
| 0 | 0 | 8395441 | 414924 |

计算出的关系对：

|  |  |  |
| --- | --- | --- |
| 关系名称 | 关系代码 | 对数 |
| 共用手机号的IMEI | 08-0003-0001-02 | 0 |
| 共用手机号的MAC | 08-0003-0002-02 | 0 |
| 共用IMEI的手机号 | 08-0003-0003-02 | 0 |
| 共用IMEI的IMSI | 08-0003-0004-02 | 257254 |
| 共用MAC的手机号 | 08-0003-0005-02 | 0 |
| 共用MAC的IMSI | 08-0003-0006-02 | 0 |
| 共用IMSI的IMEI | 08-0003-0007-02 | 178060 |
| 共用IMSI的MAC | 08-0003-0008-02 | 0 |

# 五．结果表

|  |  |  |
| --- | --- | --- |
| 字段描述 | 字段 | 备注 |
| 日期 | STAT\_DATE |  |
| 关系类型 | REL\_TYPE | 1 要素关联 2 实体关系 |
| 关系标识 | REL\_ID | 见关系编码对照表 |
| 对象A | OBJECT*ID*A |  |
| 对象A类型 | TYPE\_A | 1 人档标识，2 车档标识，3 手机号，4 IMEI, 5 IMSI, 6 MAC |
| 对象A属性 | OBJECT*ATTR*A | 人: {"PERSON*NAME":""，"ID*TYPE":"","ID\_NUMBER":}车：{"HPHM":"","HPYS":}码: 空 |
| 对象B | OBJECT*ID*B |  |
| 对象B类型 | TYPE\_B | 1 人档标识，2 车档标识，3 手机号，4 IMEI, 5 IMSI, 6 MAC |
| 对象B属性 | OBJECT*ATTR*B | "人: {""PERSON*NAME"":""""，""ID*TYPE"":"""",""ID\_NUMBER"":}车：{""HPHM"":"""",""HPYS"":}码: 空" |
| 分数 | CORE | 1~100 (百分比) |
| 最早关联时间 | FIRST\_TIME |  |
| 最早关联设备 | FIRST\_DEVICE |  |
| 最近关联时间 | LAST\_TIME |  |
| 最近关联设备 | LAST\_DEVICE |  |
| 关联次数 | TOTAL |  |
| 关系属性 | REL\_ATTR 预留(暂时为空) |  |

* 此类关系和弱关系一样放入此结果表，然后从hive到ES的同步，统一同步到es
* 【对象A属性】，【对象B属性】这两个字段可以暂时不填，【REL\_ID】见【需要计算的关系】的关系代码