移动数据全解析[[1]](#footnote-1)

一． 概述 2

二． 连接方式 2

三． 数据表说明 2

3.1 概述 2

3.2 数据表说明 4

四． 数据字典 5

4.1 由移动直接提供的表 5

4.2 基于移动数据得到的一些结果 14

五． 数据参照 15

六． 数据统计特征 16

七． 复杂统计特征 16

# 概述

本文档记录的是实验室在2010年～2011年与中国移动研究院合作所获得的河南省济源市的移动数据以及基于该数据所得到的结论和实验结果，目的在于实现实验室数据与实验的传承。

该数据的质量比较好，接近原始数据。记录了济源境内30万人一个月的汇总数据及通话详单。本数据的原始格式为由移动方面ORACLE导出的TXT文档，原始数据保存在实验室的移动硬盘中。另有一份导入到了49服务器。

# 连接方式

本数据保存在49服务器的MySQL中。访问方式为：

mysql://162.105.71.49:3306/mobiledq

username: mobile

password: mobile

其中，以JY\_开头的表均为移动数据。

该帐号的权限较高，注意不要修改原数据。

# 数据表说明

## 3.1 概述

现有移动数据主要分四大类。

**3.1.1 原始话单数据**

第一大类是原始话单数据，该数据与移动原始数据最接近，利用此一类表可以获得每个用户轨迹方面的信息。

由于移动的隐私策略，给我们的数据中，用户的号码都是加密过的，如果未经说明，数据中USER\_ID, PHONE\_NUMBER, PHONE\_A, PHONE \_B等都是指的用户ID，即加密后的用户号码，不影响实验结果。

只有一种情况例外：在JY\_RAW\_CALLING和JY\_RAW\_CALLED中，OPPO\_NUM是未加密的。这种情况使我们有可能通过时间对加密规则进行解密，但移动在记录主叫和被叫时候时间戳会有3秒钟的误差，增大了解密的困难，也让解密后的结果难以验证。

截止到第一次创建此文档，该RAW系列原始话单数据仍未完全解密，只对一些通话频繁的电话进行了解密工作。

**3.1.2 汇总信息表**

这是最早提供给我们的一组数据，包括这这个城市中移动用户的个人信息、交往数据。值得注意的是，虽然提供的是交往数据，但其统计的还是较为细致，在很多情况下完全可以抛开原始话单数据单独使用。

**3.1.3 样本及基站**

在之前的项目中，移动提供了家庭和集团样本。在实验过程中，我们发现，家庭样本准确性较高，样本量也较大。集团样本准确率较低，质量并不是太高。

基站信息表中，同一个位置会有多个基站（多向天线），AREA\_CODE为397的是在济源市内。总共有297个不同位置的基站。

**3.1.4 基于移动数据得到的一些结果**

JY\_UPDATE表记录了每个用户的轨迹，实际上是对原始话单数据的汇总。

JY\_USER\_OD表记录了每个用户的OD信息，具体方法可参见黄文灏的论文[[2]](#footnote-2)。有些用户只能发现OD中的一个，所以允许部分记录为空。

**3.1.5 LAC和CI**

在记录中有LAC和CI。其中，CI是传统意义上的基站，即CELL\_ID，LAC是多个基站的集合，是一个大区。

## 3.2 数据表说明

凡是以JY开头的表都是济源的移动数据。其中，以JY\_RAW开头的都是原始话单数据。

**3.2.1 由移动直接提供**

|  |  |  |  |
| --- | --- | --- | --- |
| 大类 | 表名 | 意义 | 其他 |
| 原始话单数据(2010年10~11月的详单) | JY\_RAW\_CALLING | 主叫表 | 一方加密，另一方未加密 |
| JY\_RAW\_CALLED | 被叫表 | 一方加密，另一方未加密 |
| JY\_RAW\_OPEN | 开机表 |  |
| JY\_RAW\_CLOSE | 关机表 |  |
| JY\_RAW\_SEND | 发送信息表 |  |
| JY\_RAW\_RECV | 接收信息表 |  |
| JY\_RAW\_HDO | 位置切换表 | 在本LAC中切换 |
| JY\_RAW\_UPDATE | 位置更新表 | 跨LAC切换 |
| 汇总信息表 | JY\_USER\_INFO | 用户信息表 | 每个用户2010年9~10月的汇总信息 |
| JY\_USER\_FINGER | 交往表 | 两个用户这一个月间 |
| 基站经纬度 | JY\_JIZHAN\_INFO | 基站表 | 基站经纬度与编号对比，其中编号是十六进制，在其他表中使用的是十进制，需要注意转换。397为济源 |
| 家庭样本 | JY\_HOME\_SERVER | 家庭信息表 | 移动客户经理提供的家庭调查表 |
| 集团样本 | JY\_GROUP\_INFO | 集团信息表 | 移动提供的集团信息 |

**3.2.2 基于移动数据得到的一些结果**

|  |  |  |  |
| --- | --- | --- | --- |
| 大类 | 表名 | 意义 | 其他 |
| 轨迹表 | JY\_UPDATE\_ALL | 轨迹表 | 汇总了原始话单数据中有关于基站切换的内容得到的衍生表 |
| 用户OD | JY\_USER\_OD | 用户OD表 | 提供了每个用户的OD经纬度信息 |

# 数据字典

在本节中，将列出上一节中所有的表结构。

## 4.1 由移动直接提供的表

**4.1.1 主叫表：JY\_RAW\_CALLING**

有多个对端号码，利用OPPO\_NUM较好。

表行数：40860678

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段 | 数据类型 | 中文含义 |
| 1 | USERID | VARCHAR2(30) | 用户id |
| 2 | CAUSE | NUMBER | 事件cause值 |
| 3 | IMSI | CHAR(18) | IMSI |
| 4 | IMEI | CHAR(18) | IMEI号 |
| 5 | TIME\_STAMP | DATE | 发生时间 |
| 6 | START\_LAC | NUMBER | 事件发起的LAC号 |
| 7 | START\_CI | NUMBER | 事件发起的CI号 |
| 8 | END\_LAC | NUMBER | 事件终止的LAC号 |
| 9 | END\_CI | NUMBER | 事件终止的CI号 |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | OPP\_PRENUM | CHAR(8) | 对端号码前缀 |
| 13 | OPP\_COUNTRY | CHAR(6) | 对端国家区号 |
| 14 | OPP\_SERV\_NUMBER | CHAR(24) | 对端号码 |
| 15 | TRAVEL\_STATE | CHAR(2) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 16 | REMOTE\_STATE | CHAR(2) | 长途状态a0:无法识别 00:本地通话  01:省内通话 02:省际通话 03:港澳台通话 04:国际通话 |
| 17 | RED\_USR | CHAR(1) | 红名单标志 |
| 18 | GRAY\_USR | CHAR(1) | 灰名单标志 |
| 19 | BLACK\_USR | CHAR(1) | 黑名单标志 |
| 20 | WHITE\_USR | CHAR(1) | 白名单标志 |
| 21 | OPP\_AREA\_CODE | CHAR(4) | 对端区号 |
| 22 | EVENT\_ID | NUMBER | 事件值20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 23 | BRAND\_ID | CHAR(1) | 品牌 |
| 24 | LOCAL\_AREA\_CODE | CHAR(4) | 本端归属地 |
| 25 | OPPO\_NUM | CHAR(32) | 原始对端号码 |
| 26 | AREA\_CODE | CHAR(4) | 地区 |
| 27 | VLR\_LOCAL | CHAR(4) | 事件发生地 |

**4.1.2 被叫表：JY\_RAW\_CALLED**

表行数：44237085

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段 | 数据类型 | 中文含义 |
| 1 | USERID | VARCHAR2(30) | 用户id |
| 2 | CAUSE | NUMBER | 事件cause值 |
| 3 | IMSI | CHAR(18) | IMSI |
| 4 | IMEI | CHAR(18) | IMEI号 |
| 5 | TIME\_STAMP | DATE | 发生时间 |
| 6 | START\_LAC | NUMBER | 事件发起的LAC号 |
| 7 | START\_CI | NUMBER | 事件发起的CI号 |
| 8 | END\_LAC | NUMBER | 事件终止的LAC号 |
| 9 | END\_CI | NUMBER | 事件终止的CI号 |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | OPP\_COUNTRY | CHAR(6) | 对端国家区号 |
| 13 | OPP\_SERV\_NUMBER | CHAR(24) | 对端号码 |
| 14 | TRAVEL\_STATE | CHAR(3) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 15 | RED\_USR | CHAR(1) | 红名单标志 |
| 16 | GRAY\_USR | CHAR(1) | 灰名单标志 |
| 17 | BLACK\_USR | CHAR(1) | 黑名单标志 |
| 18 | WHITE\_USR | CHAR(1) | 白名单标志 |
| 19 | OPP\_AREA\_CODE | CHAR(4) | 对端区号 |
| 20 | EVENT\_ID | NUMBER | 事件值20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 21 | BRAND\_ID | CHAR(1) | 品牌 |
| 22 | OPPO\_NUM | CHAR(32) | 原始对端号码 |
| 23 | LOCAL\_AREA\_CODE | CHAR(4) | 本端归属地 |
| 24 | AREA\_CODE | CHAR(4) | 地区 |
| 25 | VLR\_LOCAL | CHAR(4) | 事件发生地 |

**4.1.3 开机表：JY\_RAW\_OPEN**

表行数：4182606

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段名 | 数据类型 | 中文说明 |
| 1 | USER\_ID | CHAR(16) | 用户id |
| 2 | CAUSE | NUMBER | 事件结果CAUSE值 |
| 3 | IMSI | CHAR(18) | 用户的IMSI号 |
| 4 | IMEI | CHAR(18) | 用户IMEI号 |
| 5 | TIME\_STAMP | DATE | 事件发生时间 |
| 6 | START\_LAC | NUMBER | 业务发起的LAC编号 |
| 7 | START\_CI | NUMBER | 业务发起的CELL\_ID |
| 8 | END\_LAC | NUMBER | 业务终止的LAC编号 |
| 9 | END\_CI | NUMBER | 业务终止的CELL\_ID |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | EVENT\_ID | NUMBER | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 13 | TRAVEL\_STATE | CHAR(2) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 14 | BRAND\_ID | CHAR(1) | 用户品牌 |
| 15 | LOCAL\_AREA\_CODE | CHAR(4) | 用户归属地 |
| 16 | AREA\_CODE | CHAR(4) | 用户区号 |
| 17 | VLR\_LOCAL | VARCHAR2(4) | 用户拜访地区号 |

**4.1.4 关机表 ：JY\_RAW\_CLOSE**

表行数：3327751

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段名 | 数据类型 | 中文说明 |
| 1 | USER\_ID | CHAR(16) | 用户id |
| 2 | CAUSE | NUMBER | 事件结果CAUSE值 |
| 3 | IMSI | CHAR(18) | 用户的IMSI号 |
| 4 | IMEI | CHAR(18) | 用户IMEI号 |
| 5 | TIME\_STAMP | DATE | 事件发生时间 |
| 6 | START\_LAC | NUMBER | 业务发起的LAC编号 |
| 7 | START\_CI | NUMBER | 业务发起的CELL\_ID |
| 8 | END\_LAC | NUMBER | 业务终止的LAC编号 |
| 9 | END\_CI | NUMBER | 业务终止的CELL\_ID |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | EVENT\_ID | NUMBER | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 13 | TRAVEL\_STATE | CHAR(2) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 14 | BRAND\_ID | CHAR(1) | 用户品牌 |
| 15 | LOCAL\_AREA\_CODE | CHAR(4) | 用户归属地 |
| 16 | AREA\_CODE | CHAR(4) | 用户区号 |
| 17 | VLR\_LOCAL | VARCHAR2(4) | 用户拜访地区号 |

**4.1.5 发送信息表：JY\_RAW\_SEND**

表行数：17219530

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段名 | 数据类型 | 中文说明 |
| 1 | USER\_ID | CHAR(16) | 用户id |
| 2 | CAUSE | NUMBER | 事件结果CAUSE值 |
| 3 | IMSI | CHAR(18) | 用户的IMSI号 |
| 4 | IMEI | CHAR(18) | 用户IMEI号 |
| 5 | TIME\_STAMP | DATE | 事件发生时间 |
| 6 | START\_LAC | NUMBER | 业务发起的LAC编号 |
| 7 | START\_CI | NUMBER | 业务发起的CELL\_ID |
| 8 | END\_LAC | NUMBER | 业务终止的LAC编号 |
| 9 | END\_CI | NUMBER | 业务终止的CELL\_ID |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | EVENT\_ID | NUMBER | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 13 | OPP\_PRENUM | CHAR(18) | 对端号码前缀 |
| 14 | OPP\_COUNTRY | CHAR(4) | 用户品牌 |
| 15 | OPP\_SERV\_NUMBER | CHAR(12) | 对端号码 |
| 16 | TRAVEL\_STATE | CHAR(2) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 17 | OPPO\_NUM | CHAR(32) | 对端号码(原始信息) |
| 18 | BRAND\_ID | CHAR(1) | 品牌 |
| 19 | LOCAL\_AREA\_CODE | CHAR(4) | 本端归属地 |
| 20 | AREA\_CODE | CHAR(4) | 本端区号 |
| 21 | OPP\_AREA\_CODE | CHAR(4) | 对端区号 |
| 22 | VLR\_LOCAL | VARCHAR2(4) | 用户拜访地区号 |

**4.1.6 接收信息表：JY\_RAW\_RECV**

表行数：30694790

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段名 | 数据类型 | 中文说明 |
| 1 | USER\_ID | string | 用户id |
| 2 | CAUSE | integer | 事件结果CAUSE值 |
| 3 | IMSI | string | 用户的IMSI号 |
| 4 | IMEI | string | 用户IMEI号 |
| 5 | TIME\_STAMP | datetime | 事件发生时间 |
| 6 | START\_LAC | integer | 业务发起的LAC编号 |
| 7 | START\_CI | integer | 业务发起的CELL\_ID |
| 8 | END\_LAC | integer | 业务终止的LAC编号 |
| 9 | END\_CI | integer | 业务终止的CELL\_ID |
| 10 | HLR | string | 用户归属地信息 |
| 11 | VLR | string | 用户拜访地信息 |
| 12 | EVENT\_ID | integer | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 13 | OPP\_COUNTRY | string | 对端国家号码 |
| 14 | OPP\_SERV\_NUMBER | string | 对端号码 |
| 15 | TRAVEL\_STATE | string | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 16 | OPPO\_NUM | string | 对端号码(原始信息) |
| 17 | BRAND\_ID | string | 品牌 |
| 18 | LOCAL\_AREA\_CODE | string | 本端归属地 |
| 19 | AREA\_CODE | string | 本端区号 |
| 20 | OPP\_AREA\_CODE | string | 对端区号 |
| 21 | VLR\_LOCAL | string | 用户拜访地区号 |

**4.1.7 位置切换表：JY\_RAW\_HDO**

表行数：3565661

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段名 | 数据类型 | 中文说明 |
| 1 | USER\_ID | CHAR(16) | 用户id |
| 2 | CAUSE | NUMBER | 事件结果CAUSE值 |
| 3 | IMSI | CHAR(18) | 用户的IMSI号 |
| 4 | IMEI | CHAR(18) | 用户IMEI号 |
| 5 | TIME\_STAMP | DATE | 事件发生时间 |
| 6 | START\_LAC | NUMBER | 业务发起的LAC编号 |
| 7 | START\_CI | NUMBER | 业务发起的CELL\_ID |
| 8 | END\_LAC | NUMBER | 业务终止的LAC编号 |
| 9 | END\_CI | NUMBER | 业务终止的CELL\_ID |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | EVENT\_ID | NUMBER | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 13 | TRAVEL\_STATE | CHAR(2) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 14 | BRAND\_ID | CHAR(1) | 品牌 |
| 15 | LOCAL\_AREA\_CODE | CHAR(4) | 本端归属地 |
| 16 | AREA\_CODE | CHAR(4) | 本端区号 |
| 17 | SOUR\_LAC | NUMBER | 源LAC |
| 18 | SOUR\_CI | NUMBER | 源CI值 |
| 19 | DEST\_LAC | NUMBER | 目的LAC |
| 20 | DEST\_CI | NUMBER | 目的CI值 |
| 21 | VLR\_LOCAL | VARCHAR2(4) | 用户拜访地区号 |

**4.1.8 位置更新表：JY\_RAW\_UPDATE**

表行数：77344130

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段名 | 数据类型 | 中文说明 |
| 1 | USERID | VARCHAR2(30) | 用户id |
| 2 | CAUSE | NUMBER | 事件结果cause值 |
| 3 | IMSI | CHAR(18) | 用户IMSI |
| 4 | IMEI | CHAR(18) | 用户IMEI |
| 5 | TIME\_STAMP | DATE | 事件发送时间 |
| 6 | START\_LAC | NUMBER | 业务发起的LAC编号 |
| 7 | START\_CI | NUMBER | 业务发起的CELL\_ID |
| 8 | END\_LAC | NUMBER | 业务终止的LAC编号 |
| 9 | END\_CI | NUMBER | 业务终止的CELL\_ID |
| 10 | HLR | CHAR(24) | 用户归属地信息 |
| 11 | VLR | CHAR(24) | 用户拜访地信息 |
| 12 | RED\_USR | CHAR(1) | 红名单标志 |
| 13 | GRAY\_USR | CHAR(1) | 灰名单标志 |
| 14 | BLACK\_USR | CHAR(1) | 黑名单标志 |
| 15 | WHITE\_USR | CHAR(1) | 白名单标志 |
| 16 | EVENT\_ID | NUMBER | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 17 | TRAVEL\_STATE | CHAR(2) | 漫游状态a0无法识别:00：省际漫游用户； 01：国内漫游用户；02：非漫游用户：03:本省漫游 |
| 18 | BRAND\_ID | CHAR(1) | 品牌 |
| 19 | LOCAL\_AREA\_CODE | CHAR(4) | 本端归属地 |
| 20 | AREA\_CODE | CHAR(4) | 本端地区 |

**4.1.9 用户信息表：JY\_USER\_INFO**

表行数：365352

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 字段 | 字段含义 | 备注 |
| 1 | bill\_id | 号码 | 用户的手机号码 |
| 2 | online\_id | 在网时长 | 月份 |
| 3 | total\_fee | 上月总费用 | 单位（分） |
| 4 | call\_duration | 通话时长 | 含主被叫，单位是秒 |
| 5 | TOTAL\_DUR | 总计费时长 | 单位分钟 |
| 6 | TOTAL\_CNT | 总通话次数 | 含主被叫 |
| 7 | OUT\_DUR | 主叫计费时长 | 主叫通话的分钟数 |
| 8 | OUT\_CNT | 主叫通话次数 | 主叫通话的次数 |
| 9 | INT\_DUR | 网内计费时长 | 移动网内通话时长，单位分钟 |
| 10 | INT\_CNT | 网内通话次数 | 移动网内通话次数，单位分钟 |
| 11 | P2P\_SMS\_UP\_CNT | 点对点短信上行通信量 | 该用户发送短信的次数 |
| 12 | P2P\_SMS\_INNER\_DOWN\_CNT | 点对点短信网内下行通信量 | 该用户接受移动网内短信的次数 |
| 13 | P2P\_SMS\_OUTER\_DOWN\_CNT | 点对点短信网间下行通信量 | 该用户接受移动网外短信的次数 |
| 14 | P2P\_SMS\_INNER\_UP\_CNT | 点对点短信网内上行通信量 | 该用户发送移动网内短信的次数 |
| 15 | P2P\_SMS\_OUTER\_UP\_CNT | 点对点短信网间上行通信量 | 该用户发送移动网外短信的次数 |
| 16 | SMS\_BILL\_CNT | 短信计费量 |  |
| 17 | B\_lac\_call\_1 | 忙时最多通话基站 | 按照忙时、闲时、周末(包括节假日)统计通话最多、次多、第三代基站号、通话时长及次数。 |
| 18 | B\_lac\_cnt\_1 | 忙时最多通话基站通话次数 |
| 19 | B\_lac\_duration\_1 | 忙时最多通话基站通话时长 |
| 20 | B\_lac\_call\_2 | 忙时次多基站 |
| 21 | B\_lac\_cnt\_2 | 忙时次多通话次数 |
| 22 | B\_lac\_duration\_2 | 忙时次多通话基站通话时长 |
| 23 | B\_lac\_call\_3 | 忙时第三基站 |
| 24 | B\_lac\_cnt\_3 | 忙时第三通话次数 |
| 25 | B\_lac\_duration\_3 | 忙时第三通话基站通话时长 |
| 26 | A\_lac\_call\_1 | 闲时最多通话基站 |
| 27 | A\_lac\_cnt\_1 | 闲时最多通话基站通话次数 |
| 28 | A\_lac\_duration\_1 | 闲时最多通话基站通话时长 |
| 29 | A\_lac\_call\_2 | 闲时次多基站 |
| 30 | A\_lac\_cnt\_2 | 闲时次多通话次数 |
| 31 | A\_lac\_duration\_2 | 闲时次多通话基站通话时长 |
| 32 | A\_lac\_call\_3 | 闲时第三基站 |
| 33 | A\_lac\_cnt\_3 | 闲时第三通话次数 |
| 34 | A\_lac\_duration\_3 | 闲时第三通话基站通话时长 |
| 35 | W\_lac\_call\_1 | 周末最多通话基站 |
| 36 | W\_lac\_cnt\_1 | 周末最多通话基站通话次数 |
| 37 | W\_lac\_duration\_1 | 周末最多通话基站通话时长 |
| 38 | W\_lac\_call\_2 | 周末次多基站 |
| 39 | W\_lac\_cnt\_2 | 周末次多通话次数 |
| 40 | W\_lac\_duration\_2 | 周末次多通话基站通话时长 |
| 41 | W\_lac\_call\_3 | 周末第三基站 |
| 42 | W\_lac\_cnt\_3 | 周末第三通话次数 |
| 43 | W\_lac\_duration\_3 | 周末第三通话基站通话时长 |

**4.1.10 交往表：JY\_USER\_FINGER**

表行数：16036033

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 含义 |
| 1 | bill\_id | 用户手机号 |
| 2 | opp\_number | 对端号码 |
| 3 | opp\_type | 对端类型 |
| 4 | call\_all\_cnt | 交往总次数 |
| 5 | call\_all\_duration | 交往总时长 |
| 6 | call\_busy\_cnt | 忙时总次数 |
| 7 | call\_busy\_duration | 忙时总时长 |
| 8 | call\_free\_cnt | 闲时总次数 |
| 9 | call\_free\_duration | 闲时总时长 |
| 10 | call\_weekend\_cnt | 周末总次数482362 |
| 11 | call\_weekend\_duration | 周末总时长 |
| 12 | call\_busy\_out\_cnt | 忙时呼出次数 |
| 13 | call\_busi\_out\_duration | 忙时呼出时长（分钟） |
| 14 | call\_free\_out\_cnt | 闲时呼出次数 |
| 15 | call\_free\_out\_duration | 闲时呼出时长（分钟） |
| 16 | call\_weekend\_out\_cnt | 周末呼出次数 |
| 17 | call\_weekend\_out\_duration | 周末呼出时长（分钟） |
| 18 | call\_busy\_in\_cnt | 忙时呼入次数 |
| 19 | call\_busi\_in\_duration | 忙时呼入时长（分钟） |
| 20 | call\_free\_in\_cnt | 闲时呼入次数 |
| 21 | call\_free\_in\_duration | 闲时呼入时长（分钟） |
| 22 | call\_weekend\_in\_cnt | 周末呼入次数 |
| 23 | call\_weekend\_in\_duration | 周末呼入时长（分钟） |
| 24 | sms\_all\_cnt | 交往短信总条数 |
| 25 | sms\_busy\_cnt | 忙时短信总条数 |
| 26 | sms\_free\_cnt | 闲时短信总条数 |
| 27 | sms\_weekend\_cnt | 周末短信总条数 |
| 28 | sms\_busy\_out\_cnt | 忙时短信上行条数 |
| 29 | sms\_free\_out\_cnt | 闲时短信上行条数 |
| 30 | sms\_weekend\_out\_cnt | 周末短信上行条数 |
| 31 | sms\_busy\_in\_cnt | 忙时短信下行条数 |
| 32 | sms\_free\_in\_cnt | 闲时短信下行条数 |
| 33 | sms\_weekend\_in\_cnt | 周末短信下行条数 |

**4.1.11 基站表：JY\_JIZHAN\_INFO**

表行数：79896

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 含义 |
| 1 | bts\_key | 基站ID |
| 2 | bts\_jingdu | 基站经度 |
| 3 | bts\_weidu | 基站纬度 |

**4.1.12 家庭信息表：JY\_HOME\_SERVER**

表行数：115699

手机号码以mobile\_id为准。

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 含义 |
| 1 | home\_id | 家庭用户编号 |
| 2 | Prod\_id | 家庭产品类型 1 移动固话 2 铁通固话 |
| 3 | Phone\_num | 固话号码 |
| 4 | mobile\_id | 手机号码 |
| 5 | Role | 家庭用户角色 |

**4.1.13 集团信息表：JY\_GROUP\_INFO**

表行数：92747

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 含义 |
| 1 | group\_id | 集团ID |
| 2 | user\_id | 用户ID |

## 4.2 基于移动数据得到的一些结果

4.2.1 轨迹表：JY\_UPDATE\_ALL

表行数：160098800

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 含义 |
| 1 | BILL\_ID | 手机号码 |
| 2 | START\_LAC | 开始的LAC |
| 3 | START\_CELL\_ID | 开始的基站编号 |
| 4 | END\_LAC | 结束的LAC |
| 5 | END\_CELL\_ID | 结束的基站编号 |
| 6 | EVENT\_TIME | 时间发生的时间爱你 |
| 7 | EVENT\_TYPE | 事件编号20: 主叫21：被叫50：开机51：关机40: 短信发送41：短信接收30:位置更新31：位置切换 |
| 8 | AREA\_CODE | 区域，397 |

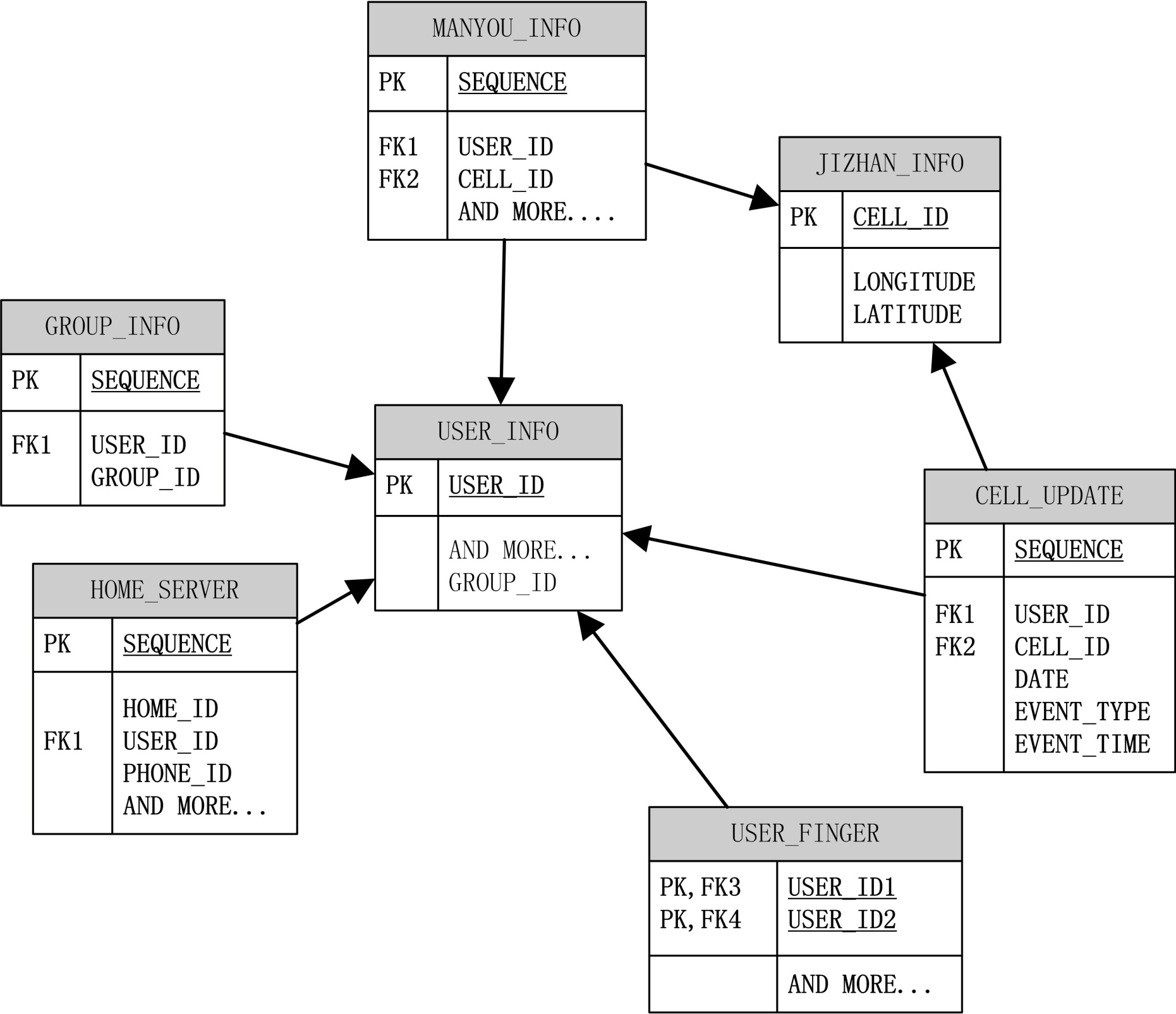
4.2.2 用户OD表

表行数：343404

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 含义 |
| 1 | PHONE\_NO | 手机号码 |
| 2 | OCELL\_ID | 用户O的基站编号 |
| 3 | OCELL\_LONGITUDE | 用户O的经度 |
| 4 | OCELL\_LATITUDE | 用户O的纬度 |
| 5 | DCELL\_ID | 用户D的基站编号 |
| 6 | DCELL\_LONGITUDE | 用户D的经度 |
| 7 | DCELL\_LATITUDE | 用户D的纬度 |

# 数据参照

移动数据的结构并不复杂，并且基本满足参照性，见下图（未列出原始话单数据表）。



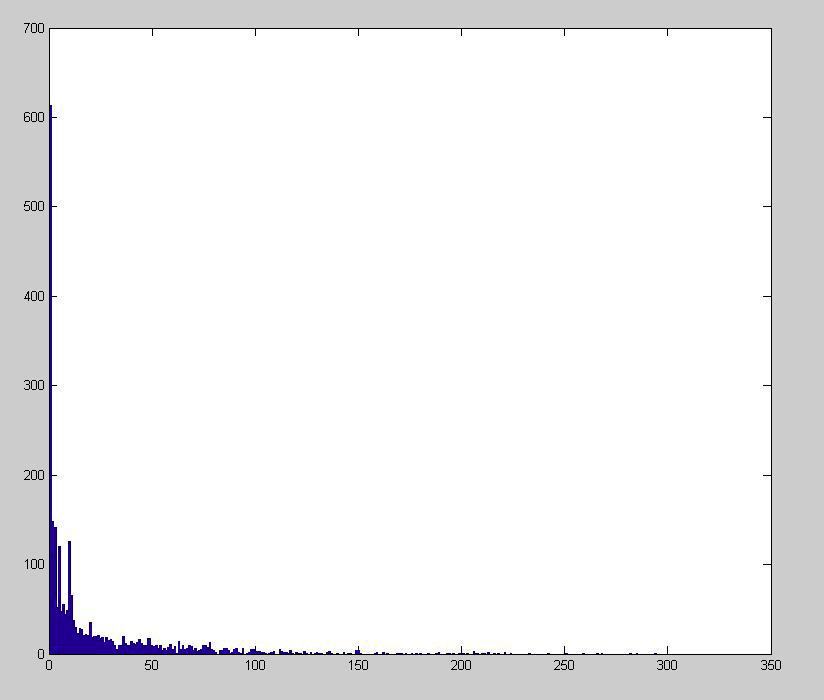
# 数据统计特征

USER\_FINGER:

平均一个人有82个联系人

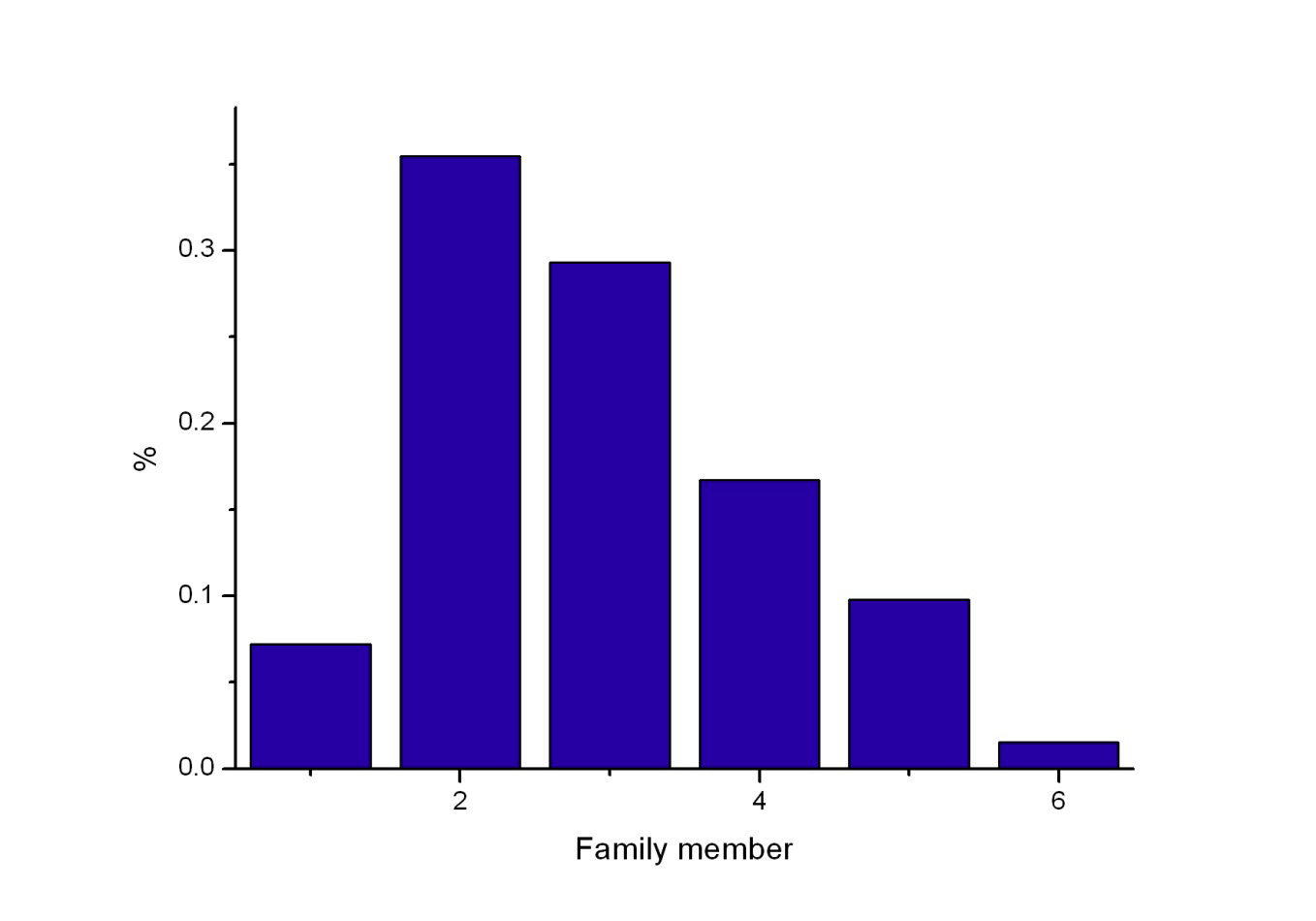
GROUP\_INFO:

有2537个group，平均一个集团36.58个人，最大的集团5699个人，编号为HJ04J251175。分布为：



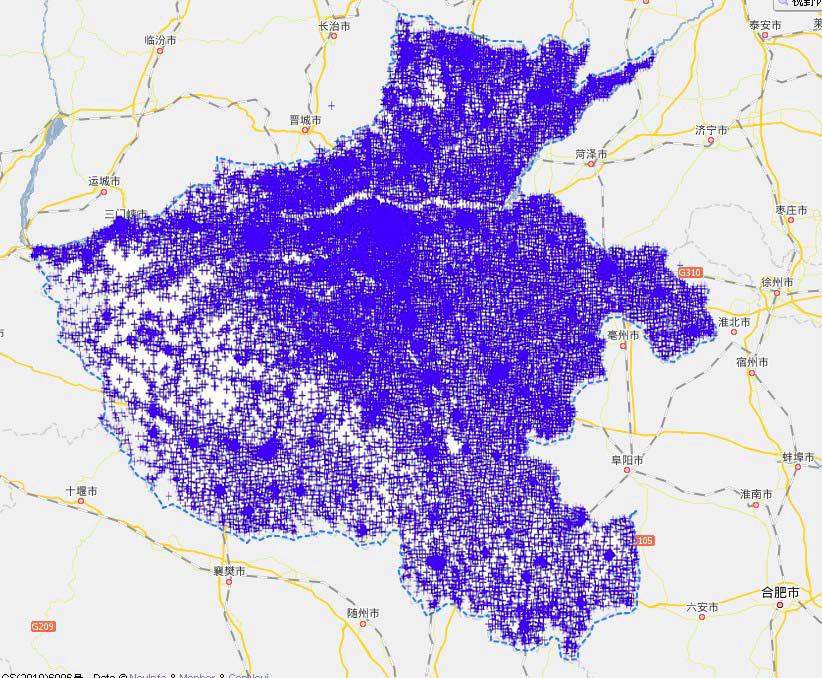
HOME\_SERVER:

有39749个home，平均一个家庭2.91个人。分布为：



JIZHAN\_INFO:

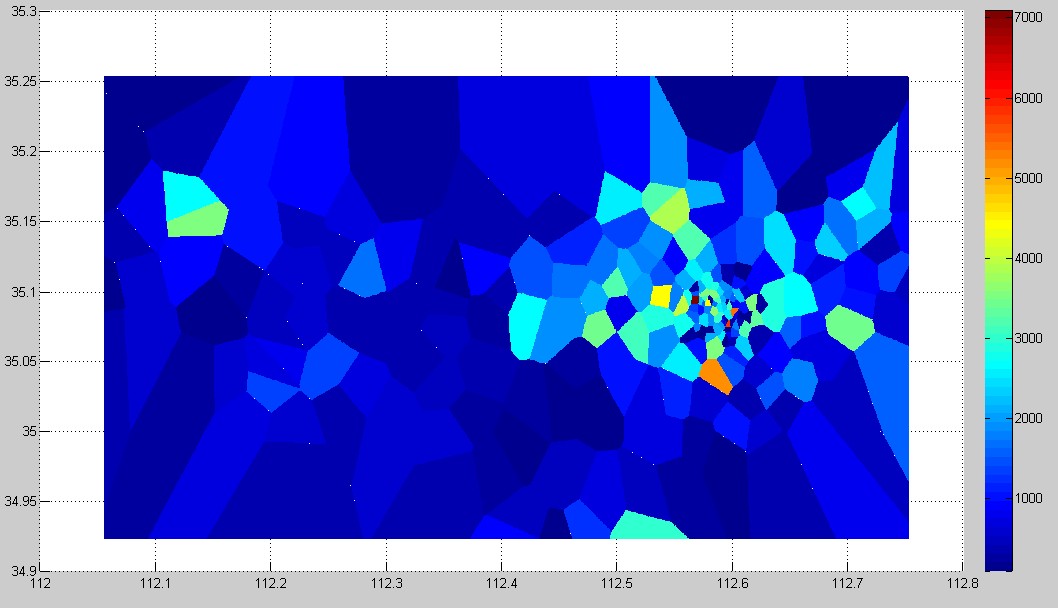
CELL\_ID是十六进制的，在地图上分布如下。说明所给的其实是河南全境的基站数据。



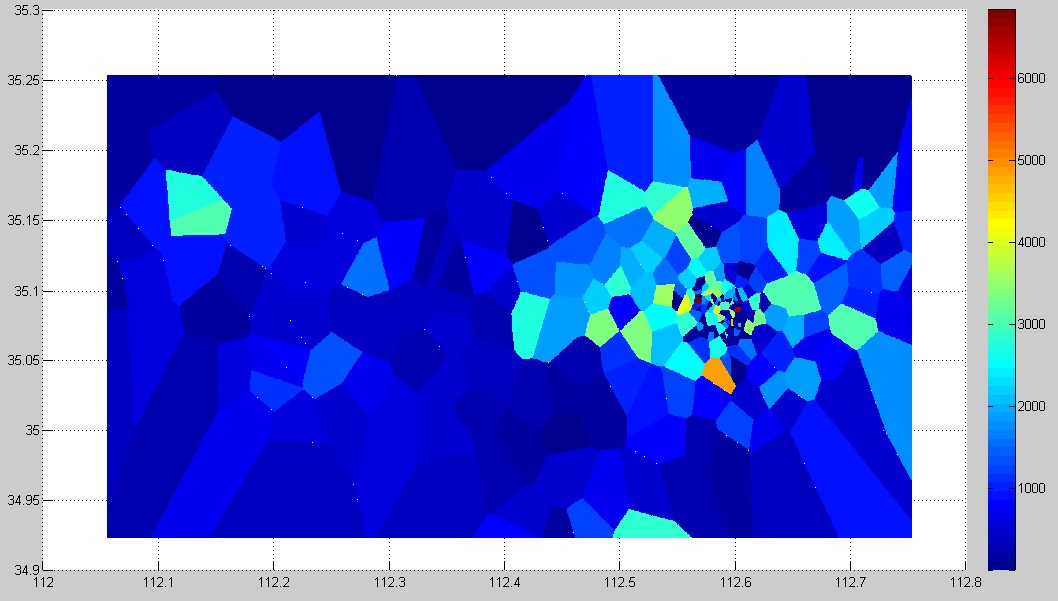
USER\_OD:

济源市OD的分布如图：

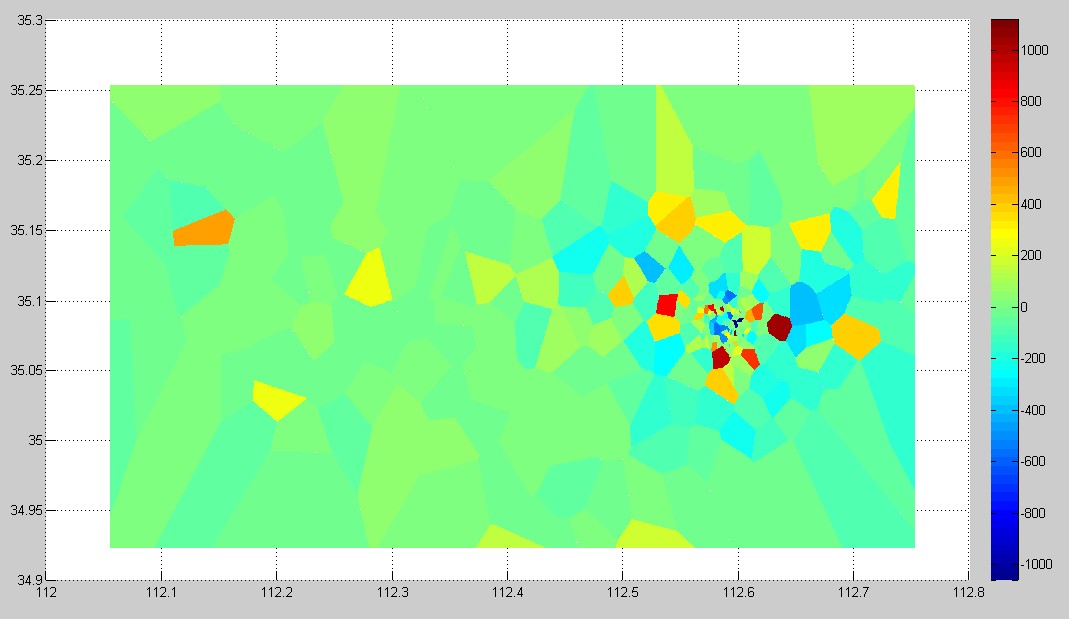
D的分布：



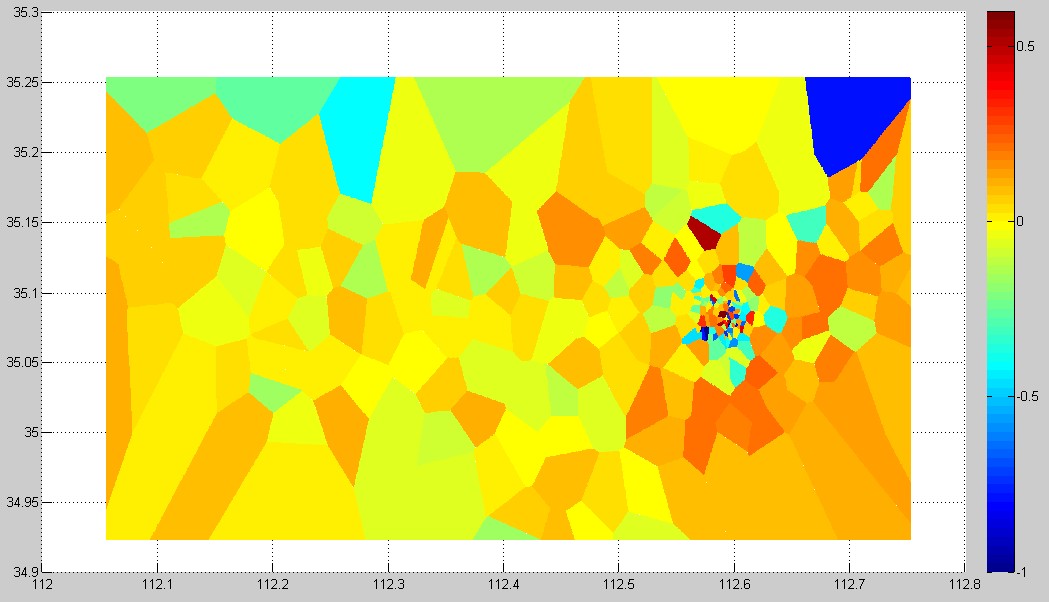
O的分布：



D-O的分布：

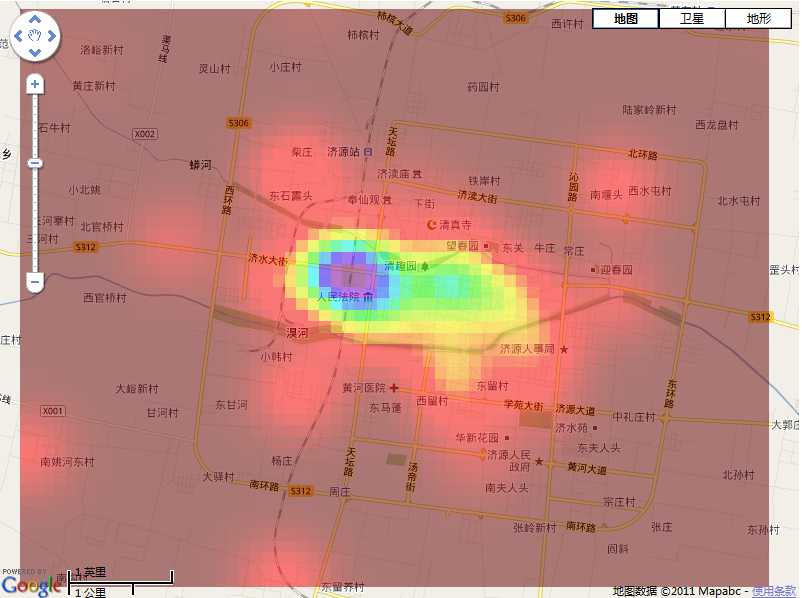


绝对值

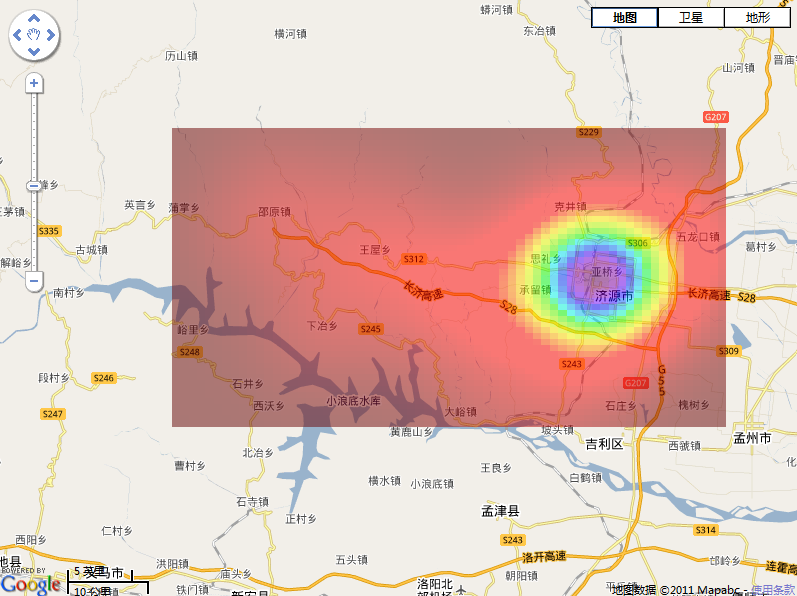


比例

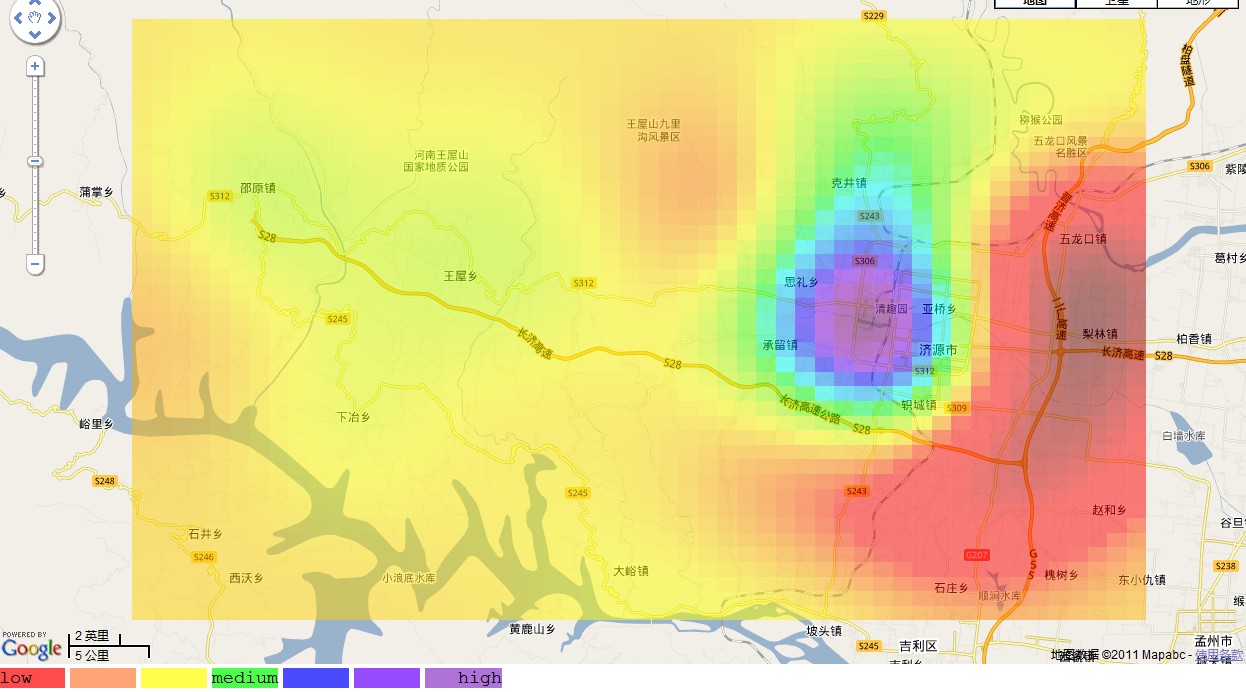
城市中心人群热图：



济源市人群分布热图：



商业居民区分布热图：



# 复杂统计特征

1. 2011.10.8 ver0.1 created by ZhaoNan [↑](#footnote-ref-1)
2. 1 Anchor Points Seeking of Large Urban Crowd Based On the Mobile Billing Data [↑](#footnote-ref-2)