选取处理好的一个月数据进行作图，实现数据的可视化，总体分析不同数据对模型预判的贡献度。

漫游地区号和对端号码归属区号的异常率分布起伏较大，不同客户的位置移动频率是不同的，有很大差别。但大多数用户的位置信息异常率为0，因此异常率对于异常率不为0的用户识别贡献度很高，但对异常率为0的用户识别贡献率为0.

由于不同用户的工作需要和交通圈大小不同，其月通话次数也会有很大不同。从其分布图来看，不同用户的约通话次数分布变化很大，对识别不同的用户贡献度较高。

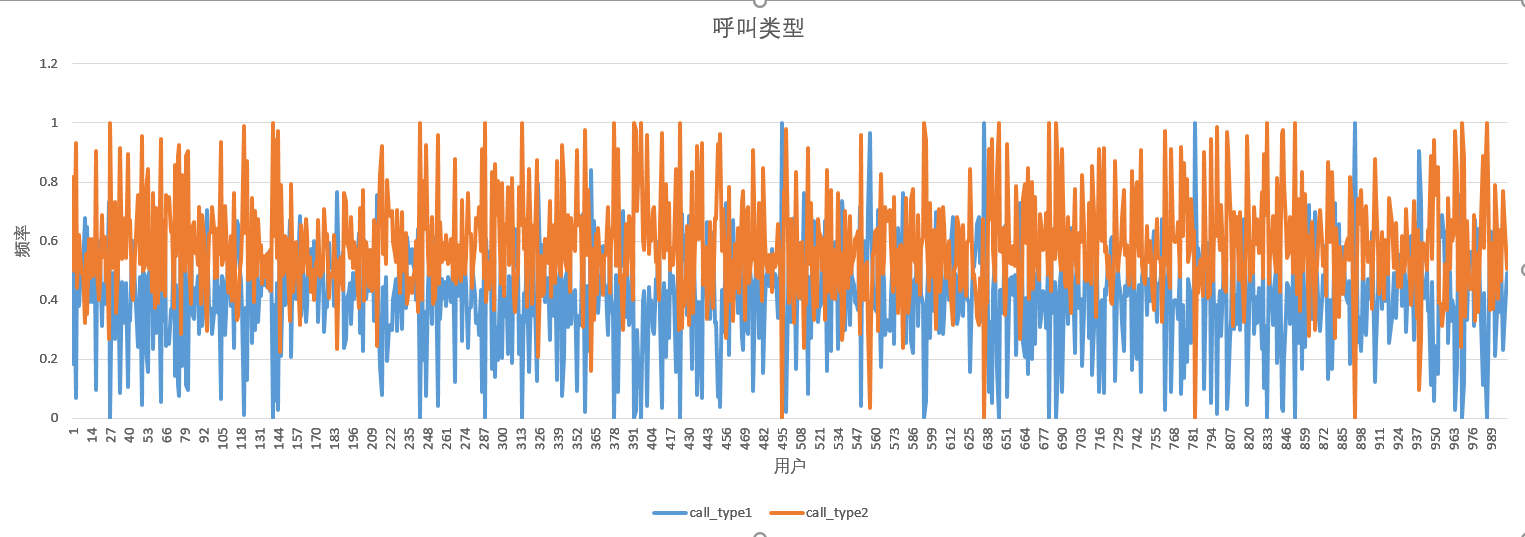
以四小时为单位将一天分为六个时段。可以发现每个用户在不同时间段的通话次数是不同的，在六个时段上大体上均呈递增分布。同时每个用户在一天各个时段通话次数的分布是不同的，有的在六时段分布比较平均，有的则随六时段有很大的变化。这反应了用户每天的作息规律，有较大的互异性。六时段分布随用户不同有很大不同，其模型贡献度将比较大。

以上中下旬为分类标准统计用户一个月不同时间的通话次数。可以发现大多数用户三个时期的通话次数差别不大，只有个别用户会有较大区别，可以反应用户有可能从事的特殊工作类型。对模型的贡献度要低于六时段分布。

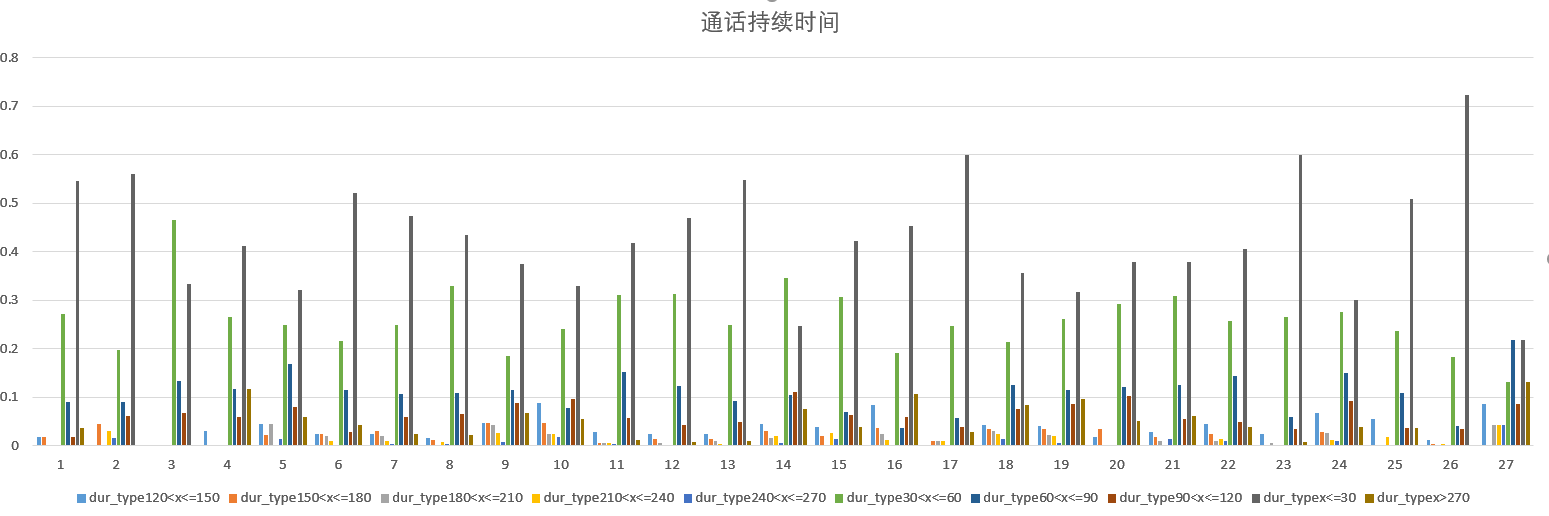
如图提取了部分用户与对端号码相互呼叫频次前5位的数值，从该图可以看出大部分用户均保持着一种通话规律，即存在频繁通话对象和一般通话对象。活跃用户的频繁通话对象群较多，即超过5个，非活跃用户的频繁通话对象群数量较低，这个可以作为一个用户交往群的重要特征。

如图提取了部分用户Lac基站出现频次排名前3位的数值，从该图可以看出大部分用户基站信息相对保持稳定。活跃用户的基站频次显示出固定的特征，部分用户存在不同基站的情况。这也可以作为用户的通话特征采用，可以显示出不同用户的通话区域特性。

如图提取了部分用户Cellid通话位置信息频次前3位的数值，从该图可以看出大部分活跃用户的位置信息比较稳定，即“一枝独秀”的情况比较明显，也存在一些活跃用户有一些不同的情况。可用这一信息来进一步区别用户之间的区别，并通过用户的日常通话位置信息丰富交往圈指纹。



将用户按呼叫类型分为主叫（call\_type1）和被叫（call\_type2），由上图可知，用户作为被叫的频率更高。



用户每次通话时间以30s为间隔分类，用来反映用户每次通话时间趋势。由上图可看出：在选取的用户中，通话时间基本控制在60s以内，其中绝大部分在30s以内，这可以作为用户的通话特征。

由图提取了部分用户漫游类型的频率，可以看出数据起伏很小，基本集中在longdis\_type0，即通话基本是本地非漫游，无法利用该特征区别用户，但易于提取拥有其他漫游类型的用户。

该图选取了部分用户通话对象号码的运营商类型，根据分布特点可看出，由于各用户的交友圈差异，不能得出有利的运营商趋势，但初步看出移动用户占比最大。需要进一步分析运营商特征，以得到更佳的交友圈特征。

该图选取的为部分用户的通话类型，可看出数据起伏不大，集中在roam\_type0，即用户之间为基本通话，该特征对用户分类贡献率低。