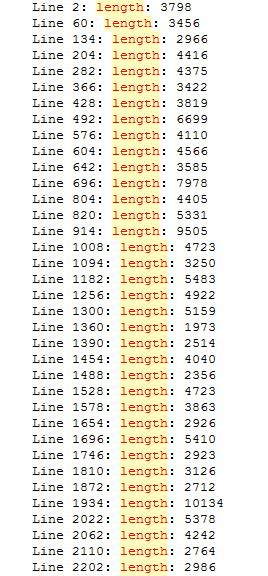
# Mall wifi



Mall shop

Line 2: shoplength: 71

Line 4: shoplength: 132

Line 6: shoplength: 114

Line 8: shoplength: 104

Line 10: shoplength: 73

Line 12: shoplength: 88

Line 14: shoplength: 109

Line 16: shoplength: 70

Line 18: shoplength: 94

Line 20: shoplength: 89

Line 22: shoplength: 76

Line 24: shoplength: 86

Line 26: shoplength: 104

Line 28: shoplength: 93

Line 30: shoplength: 81

Line 32: shoplength: 53

Line 34: shoplength: 80

Line 36: shoplength: 75

Line 38: shoplength: 76

Line 40: shoplength: 62

Line 42: shoplength: 76

Line 44: shoplength: 86

Line 46: shoplength: 68

Line 48: shoplength: 72

Line 50: shoplength: 76

Line 52: shoplength: 78

Line 54: shoplength: 59

Line 56: shoplength: 92

Line 58: shoplength: 93

Line 60: shoplength: 137

Line 62: shoplength: 128

Line 64: shoplength: 113

Line 66: shoplength: 87

Line 68: shoplength: 83

Line 70: shoplength: 89

Line 72: shoplength: 78

Line 74: shoplength: 68

Line 76: shoplength: 124

Line 78: shoplength: 220

Line 80: shoplength: 122

Line 82: shoplength: 108

Line 84: shoplength: 90

Line 86: shoplength: 73

Line 88: shoplength: 85

Line 90: shoplength: 62

Line 92: shoplength: 74

Line 94: shoplength: 97

Line 96: shoplength: 78

Line 98: shoplength: 90

Line 100: shoplength: 46

Line 102: shoplength: 71

Line 104: shoplength: 79

Line 106: shoplength: 79

Line 108: shoplength: 93

Line 110: shoplength: 94

Line 112: shoplength: 80

Line 114: shoplength: 80

Line 116: shoplength: 61

Line 118: shoplength: 83

Line 120: shoplength: 81

Line 122: shoplength: 67

Line 124: shoplength: 120

Line 126: shoplength: 75

Line 128: shoplength: 125

Line 130: shoplength: 96

Line 132: shoplength: 69

Line 134: shoplength: 71

Line 136: shoplength: 118

Line 138: shoplength: 83

Line 140: shoplength: 68

Line 142: shoplength: 58

Line 144: shoplength: 92

Line 146: shoplength: 89

Line 148: shoplength: 87

Line 150: shoplength: 123

Line 152: shoplength: 79

Line 154: shoplength: 69

Line 156: shoplength: 80

Line 158: shoplength: 74

Line 160: shoplength: 109

Line 162: shoplength: 66

Line 164: shoplength: 76

Line 166: shoplength: 95

Line 168: shoplength: 99

Line 170: shoplength: 79

Line 172: shoplength: 125

Line 174: shoplength: 58

Line 176: shoplength: 94

Line 178: shoplength: 91

Line 180: shoplength: 89

Line 182: shoplength: 74

Line 184: shoplength: 96

Line 186: shoplength: 78

Line 188: shoplength: 72

Line 190: shoplength: 77

Line 192: shoplength: 82

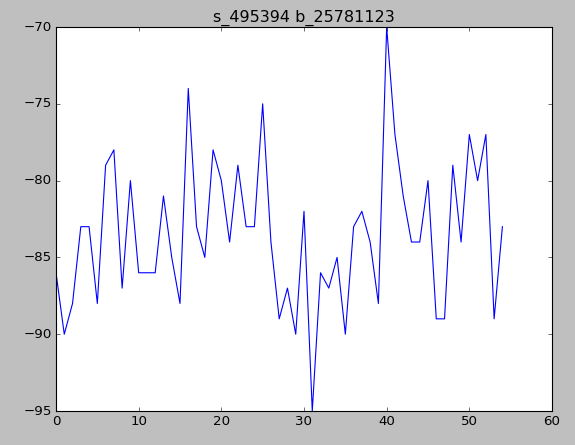
Line 194: shoplength: 91

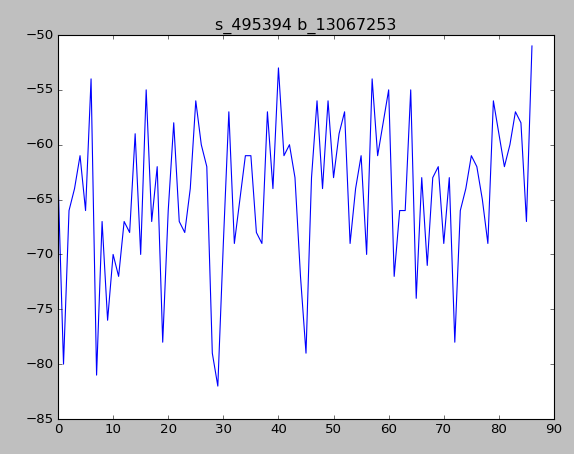
,如果我们能够通过分析找出真正 对位置判断有效的数据分量,就可以明显提高定位系统的精度.为此,本文提出了一种基于主成分分析的指纹定 位模型,通过主成分分析技术,找到不同环境条件下影响定位精度的主要分量,提高指纹的可区分度,进而提高 指纹定位的精度.同时,主成分分析也能够对指纹向量进行降维处理,提升指纹定位算法的处理效率.主成分分 析方法是从数据源优化的角度来提高定位精度,因此可以应用于大部分的指纹算法以提高精度

1. 记录含有多个相同的bssid:13019条

RSS随时间变化，而且有时变化得很显著。不过在五分钟内RSS的变化基本不会超过20dB

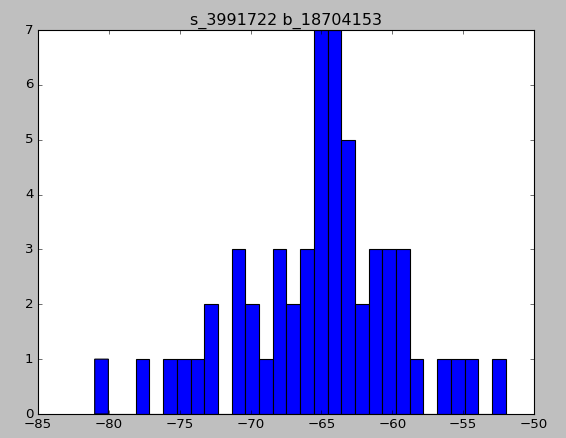
同一个商店bssid很多，有部分只出现一次，能否删除，





Shop\_id的信息看成一篇文档，bssid看成词，（关键词提取）测试集给出词，来判断该文本属于哪一类





Linux统计文件个数

ls -l |grep "^-"|wc -l

0.944

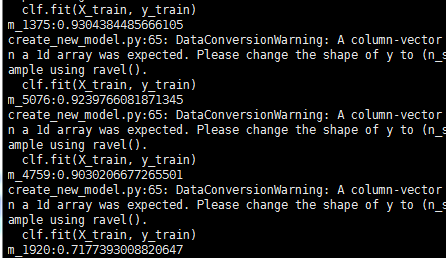
Train\_data:

user\_id,shop\_id,time\_stamp,longitude,latitude,wifi\_infos

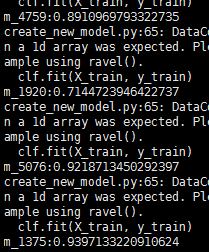
Test\_data

row\_id,user\_id,mall\_id,time\_stamp,longitude,latitude,wifi\_infos

-20



-10





Libgbm,fm,ffm

0.8782 -15，-100为默认值，没有去除缺失值（没选到bssid）样本 rf\_change\_label\_100

0.84，0为默认值

将时间因素考虑进去

周六和周日单独建模？

时间越近，建模越好？

显然，基于bagging的随机森林模型和基于boosting的Gradient Tree Boosting模型有着不少共同的参数。随机森林的子模型需要拥有较低的偏差，整体模型的训练过程旨在降低方差，记得上面说到，有两种方式，第一增加模型数量，第二减少模型之间的相关度，因此，n\_estimators默认值为10 ， 且子模型不是弱模型(max\_depth默认值设置为None，也就是说树的深度足够深)，同时，降低子模型的相关度(max\_features设置为总特征数的开方)；另一方面，GradientTreeBoosting的子模型都是有着较低的方差，整个模型的目的是为了降低偏差，因此，需要较多的子模型(n\_estimators)且子模型为弱模型(max\_depth默认为3)，但是降低子模型之间的相关度并不会显著减少整体模型的方差(max\_features设置为None)

0.879 rf estimator 50

0.8806,增加连接的bssid

加入扫描数特征预测很奇怪

0.881左右融合特征 取前三与特征选择20的交集

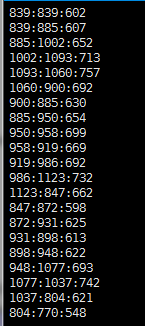
变为100个估算器 0.8829

加了经纬度0.884左右

0.8829换了连接的bssid

不加前三0.8824

换了连接的bssid,加了经纬度前五位



每一天和前一天的重合

经纬5位150estimator 0.8944

50 estimator 0.8918

中位数 0.8914

修改强度距离计算 0.89498

按距离找特征 0.8979 estimatore 150

Add 经纬度 0.89833

Importance是？

Add 新的bssid 0.8990,商场中前几个比较大的

Mallbssiddic取的shop\_impor\_dic[bssid]=len(strengths)/-np.max(str)

取最多的shopid与最小的shopid 的经纬度计算距离

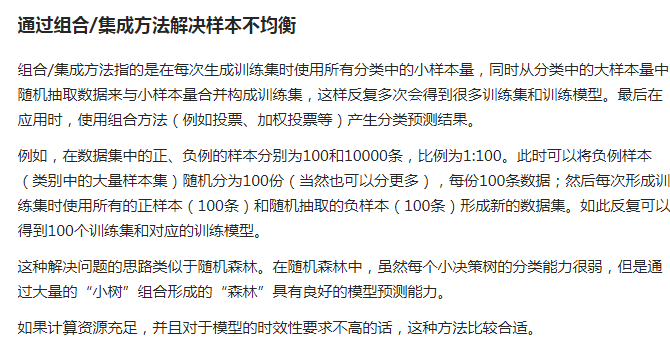
0.79118

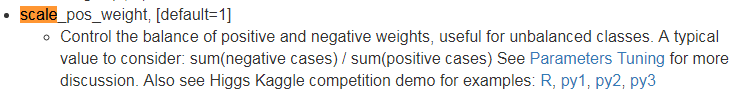


Len改为bssidlen

0.71,decisiontree加入randomfrorest变为0.73，加xg 0.75,全部特征0.76

0.7930 randomfroest加入xgboost 0.757





<https://zhuanlan.zhihu.com/p/24877060>

概率较大的形成簇，0.01用的每个shop 特征rf 0.788正确率 用部分特征0.782

Xg 0.798 部分特征 0.7868

大概率作为小概率负样本，0.20，randomforest全部特征