# 人体状态预测

可穿戴设备可以使我们有效的利用传感器获取人类的生理数据，本项目的目标是根据传感器收集到的数据对人体的坐姿进行基本分析。

现在我们收集了A，B，C，D，E一共五位用户的可穿戴设备上的传感器数据，每位用户的数据集包含一个特征文件和一个标签文件。

特征文件中每一行对应一个时刻的所有传感器数值，标签文件中每行记录了 和特征文件中对应时刻的标记过的用户姿态，两个文件的行数相同，相同行之间互相对应。

特征文件一共包含41列数据：第一列为时间戳，第二列心率。后续30多行的数据是由三个传感器依次记录的数据：

在传感器1对应的13列数据特征中，包含：1项温度数据、3项一型三轴加速 度数据、3项二型三轴加速度数据、3项三轴陀螺仪数据和3项三轴磁场数据。

人体的温度数据可以反映当前活动的剧烈程度，一般在静止状态时，体温趋 于稳定在36.5度上下；当温度高于37度时，可能是进行短时间的剧烈运动，比 如跑步和骑行。

在数据中有两个型号的加速度传感器，可以通过互相印证的方式，保证数据 的完整性和准确性。通过加速度传感器对应的三个数值，可以知道空间中x、y、 z三个轴上对应的加速度，而空间上的加速度和用户的姿态有密切的关系，比如用户向上起跳时，z轴上的加速度会激增。

陀螺仪是角运动检测的常用仪器，可以判断出用户佩戴传感器时的身体角度是水平、倾斜还是垂直。直观地，通过这些数值都是推断姿态的重要指标。

磁场传感器可以检测用户周围的磁场强度和数值大小，这些数据可以帮助我们理解用户所处的环境。比如在一个办公场所，用户座位附近的磁场是大体上固定的，当磁场发生改变时，我们可以推断用户的位置和场景发生了变化。

标签文件包含如下：每一行代表与特征文件中对应行的用户姿态类别。 总共有0-24共25种身体姿态，如，无活动状态，坐态、跑态等。标签文件作为 训练集的标准参考准则，可以进行特征的监督学习。我们可以选定某种分类模型（或者说是 算法），通过使用训练数据进行模型学习，然后对每个测试样本给出对应的分类结果。

## 算法流程

1. 读取文件，对数据进行预处理。
2. 创建分类器，对数据开展训练。
3. 计算模型准确率和召回率，评估模型。

具体实现：

1. 调用pandas的read\_table（）函数对文件进行读取。
2. 将前四个用户数据作为训练集，最后一个用户E作为测试集。故使用train\_test\_split打乱数据，将test\_size赋值为0，即打乱训练集数据即可。
3. 将训练集送入K近邻模型进行训练，并进行预测，得出其准确率、召回率、F1值。

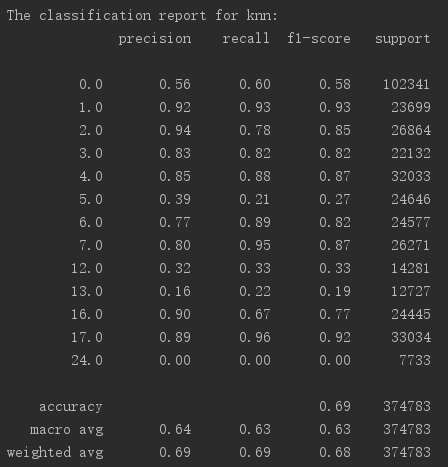
注：

**1.精确率(Precision）**是指在所有系统判定的“真”的样本中，确实是真的的占比，就是TP/(TP+FP)。

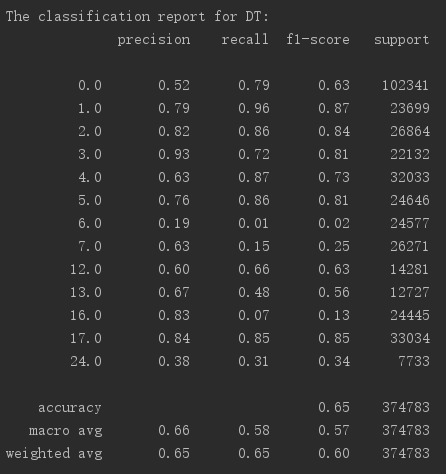
**2.召回率（Recall）**是指在所有确实为真的样本中，被判为的“真”的占比，就是TP/(TP+FN)。

3. **F1值**是为了综合考虑精确率和召回率而设计的一个指标，一般公式为取P和R的harmonic mean:2\*Precision\*Recall/(Precision + Recall)。

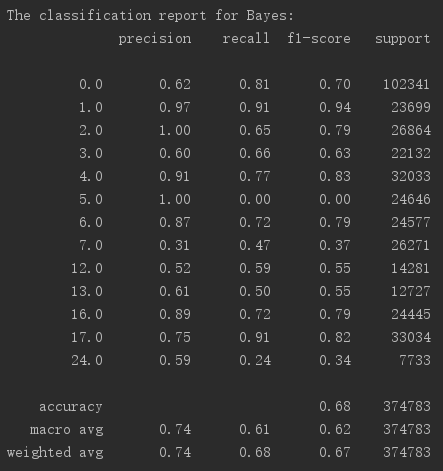
结果如下：



如果使用决策树算法，则结果如下：



如果使用贝叶斯，结果如下：



结果对比：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 准确率 | 召回率 | F1值 |
| KNN | 0.69 | 0.69 | 0.68 |
| 决策树 | 0.65 | 0.65 | 0.60 |
| 贝叶斯 | 0.74 | 0.68 | 0.67 |

结论如下：

1.从准确度的角度衡量，贝叶斯分类器的效果最好

2.从召回率和F1值的角度衡量，k近邻效果最好

3.贝叶斯分类器和k近邻的效果好于决策树