

监督学习

ML12



礼欣

www.python123.org



人体运动状态预测-实例分析

背景介绍

- 可穿戴式设备的流行，让我们可以更便利地使用传感器获取人体的各项数据，甚至生理数据。
- 当传感器采集到大量数据后，我们就可以通过对数据进行分析和建模，通过各项特征的数值进行用户状态的判断，根据用户所处的状态提供给用户更加精准、便利的服务。

数据介绍

- 我们现在收集了来自 A,B,C,D,E 5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（`a.feature`）和一个标签文件（`a.label`）。
- 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。

数据介绍-feature

我们的特征文件共包含41列特征，数据内容如下图

The image shows a data file with 41 columns. The first three columns are labeled 'column 1', 'column 2', and 'column 3'. The data is presented in rows, with the first three rows labeled 'row 1', 'row 2', and 'row 3'. The data values are numerical, and some values are missing, indicated by a question mark. A red line points to the first three columns, and another red line points to a missing value indicator. The data is presented in a table format with a header row and several data rows.

	column 1	column 2	column 3	...
row 1	8.38	104.30	2.37223	8.60074, 3.51048, 2.43954, 8.76165, 3.35465, -0.0922174, 0.0568115, -0.0158445, 14.6806, -69.212
row 2	8.39	?, 30	2.18837	8.5656, 3.66179, 2.39494, 8.55081, 3.64207, -0.0244132, 0.0477585, 0.00647434, 14.8991, -69.2224, -5.82
row 3	8.4	?, 30	2.37357	8.60107, 3.54898, 2.30514, 8.53644, 3.7328, -0.0579761, 0.0325743, -0.00698815, 14.242, -69.5197, -5.12
...	8.41	?, 30	2.07473	8.52853, 3.66021, 2.33528, 8.53622, 3.73277, -0.0023516, 0.0328098, -0.00374727, 14.8908, -69.5439, -6.1
	8.42	?, 30	2.22936	8.83122, 3.7, 2.23055, 8.59741, 3.76295, 0.0122691, 0.018305, -0.0533248, 15.5612, -68.8196, -6.28027
	8.43	?, 30	2.29959	8.82929, 3.5471, 2.26132, 8.65762, 3.77788, 0.00323826, 0.0122597, -0.0544735, 15.4565, -68.8
	8.44	?, 30	2.33738	8.829, 3.54767, 2.27703, 8.77828, 3.7323, -0.0237041, -0.0315184, -0.0478827, 15.1206, -68.57
	8.45	?, 30	2.37142	?, 0.055, 3.39347, 2.39786, 8.89814, 3.64131, 906, -0.022004, 0.033, 173, 14.8, 19, -68.730
? means a missing value	8.46	?, 30	2.3395	13251, 3.54668, 2, 37, 8.5, 241, 3.6, 02, 4, 530, -0, 1776, 769, 19, 1.9
	8.47	?, 30	2.3559	2415, 3.430, 8, 9, 1, 19, 1.9

数据介绍-feature

特征文件的各项特征具体如下表所示

1	2	3-15	16-28	29-41
时间戳	心率	传感器1	传感器2	传感器3

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。

3	4-6	7-9	10-12	13-15
温度	一型三轴加速度	二型三轴加速度	三轴陀螺仪	三轴磁场

数据介绍-feature

1	2	3-15	16-28	29-41
时间戳	心率	传感器1	传感器2	传感器3

3	4-6	7-9	10-12	13-15
温度	一型三轴加速度	二型三轴加速度	三轴陀螺仪	三轴磁场

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。

数据介绍-feature

1	2	3-15	16-28	29-41
时间戳	心率	传感器1	传感器2	传感器3

3	4-6	7-9	10-12	13-15
温度	一型三轴加速度	二型三轴加速度	三轴陀螺仪	三轴磁场

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。

数据介绍-feature

1	2	3-15	16-28	29-41
时间戳	心率	传感器1	传感器2	传感器3

3	4-6	7-9	10-12	13-15
温度	一型三轴加速度	二型三轴加速度	三轴陀螺仪	三轴磁场

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。

数据介绍-feature

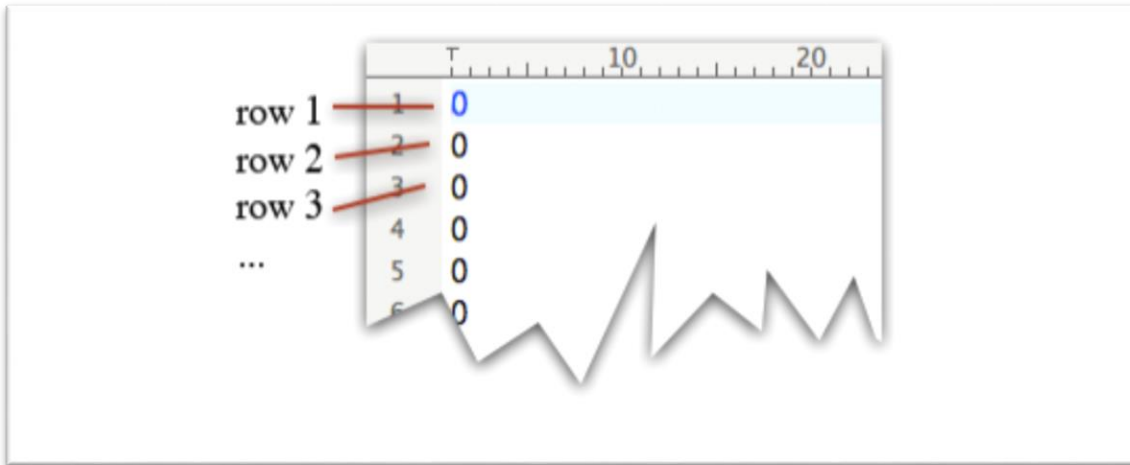
1	2	3-15	16-28	29-41
时间戳	心率	传感器1	传感器2	传感器3

3	4-6	7-9	10-12	13-15
温度	一型三轴加速度	二型三轴加速度	三轴陀螺仪	三轴磁场

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。

数据介绍-label

标签文件内容如图所示，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。



任务介绍

- 假设现在出现了一个新用户，但我们只有传感器采集的数据，那么该如何得到这个新用户的姿态呢？
- 又或者对同一用户如果传感器采集了新的数据，怎么样根据新的数据判断当前用户处于什么样的姿态呢？

任务介绍

在明确这是一个分类问题的情况下，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。

机器学习的分类算法众多，在接下来的学习中我们将会详细介绍经典的分类算法，如K近邻、决策树和朴素贝叶斯的原理和实现。