

步态识别行人分类实验



实验目的

- 综合运用之前课程中学习的各种机器学习算法，完成步行传感器数据的分类实验
- 通过完整的数据处理，算法选择，实验设计，结果分析流程掌握并深刻理解各种机器学习算法
- 具备使用机器学习算法解决实际问题的能力
- 培养科学研究的一般思路和方法

实验任务

- 对给定的数据进行预处理
- 将数据处理成为算法可以使用的格式
- 对传感器数据进行分类，判断其属于哪一个人

实验环境

- 硬件：x86_64 Centos 3.10.0 服务器/GPU 服务器、GPU、PC 上位机
- 软件：SimpleAI 实验平台、Docker 下 Ubuntu16.04 镜像、python3.5

背景

手机传感器 [1]

为满足手机的日常使用场景，现代智能手机内置了多种传感器，本实验主要介绍两种

- 加速传感器：加速度传感器，顾名思义就是一种能够测量加速度的电子设备。运用压电效应实现，重力感应模块由一片“重力块”和压电晶体组成，当手机发生动作的时候，重力块会和手机受到同一个加速度，这样重力块作用于不同方向的压电晶体上的力也会改变，这样输出的电压信号也就发生改变，根据输出电压信号就可以判断手机的方向了。
- 陀螺仪：陀螺仪是一种用于测量角度以及维持方向的设备，原理是基于角动量守恒原理。陀螺仪的用途主要是手机的摇一摇，或者在某些游戏中可以通过移动手机改变视角或VR应用。

步态

- 步态是指行人走路时所表现的姿态及走路所有的动作，这是一种复杂的行为特征，每个人都拥有一种与众不同的步态
- 步态唯一性的物理基础是每个人生理结构的差异性，不同的腿骨长度、不同的肌肉强度等共同决定了步态的唯一性
- 步态是一种非接触的生物特征识别技术，在识别的过程中不需要人的行为配合
- 步态识别是一种新兴的生物特征识别技术

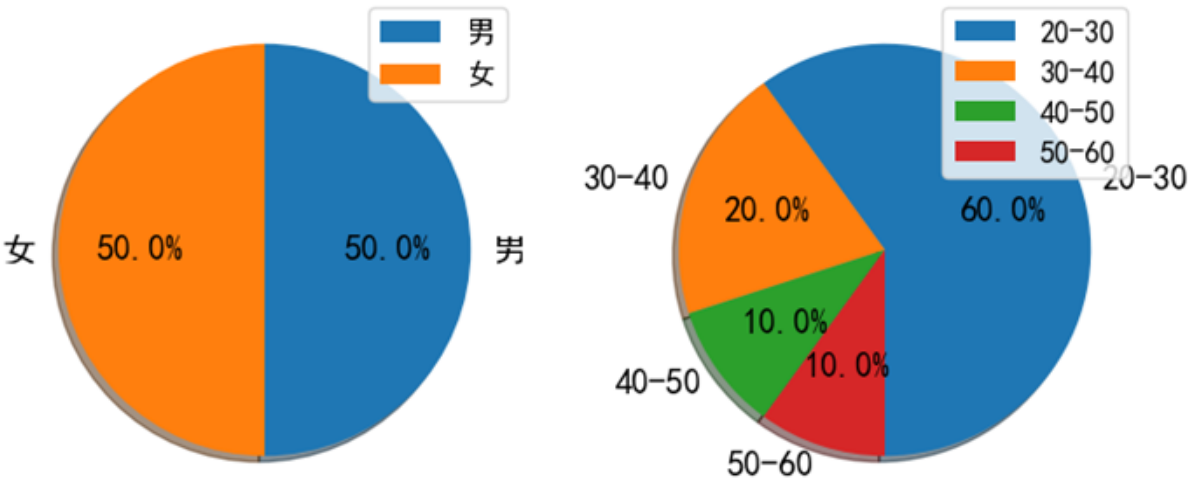
研究步态的意义

- 利用步态进行身份识别
- 利用步态进行健康检查、疾病诊断
- 利用步态进行行为预测、步行状态分析

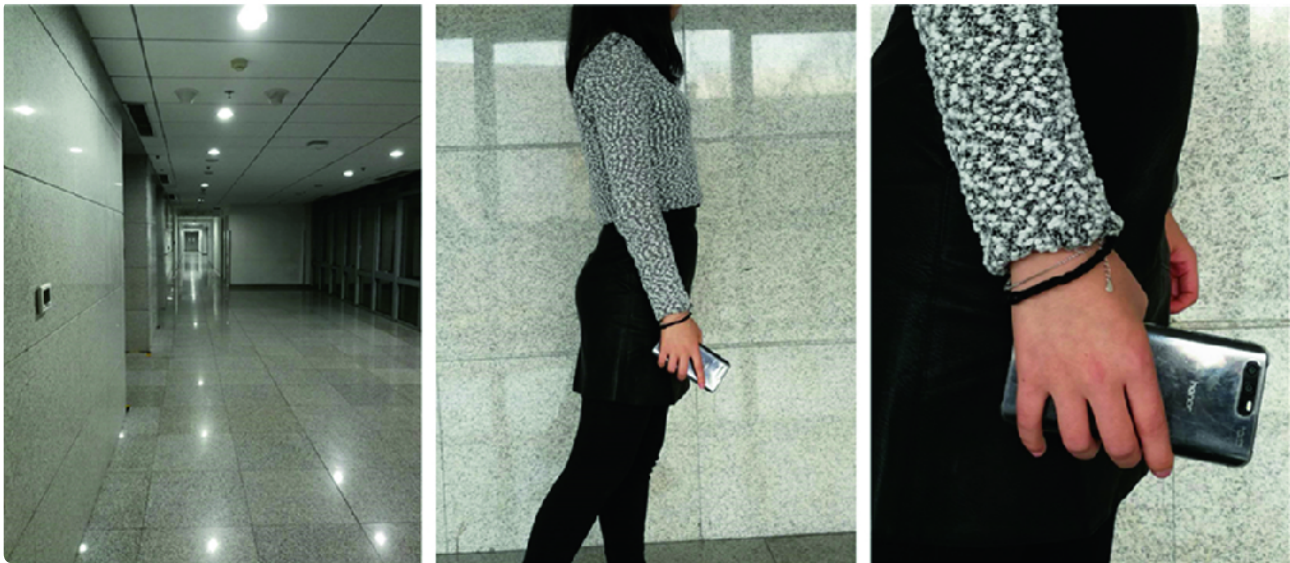
数据

数据采集信息

- 行人个数：10。行人性别、年龄分布见下图。























- 数据采集条件：行人右手持手机在自由状态下行走10分钟




- 数据格式：时间戳 X轴数据 Y轴数据 Z轴数据
- 采样频率：50HZ，即每20ms采样一次
- 数据收集时间：2017-10-26

数据详情

数据在文件夹 `data` 中，`data` 中共有20个文件，分别为 `accData0.txt` ~ `accData9.txt` 和 `gyrData0.txt` ~ `gyrData9.txt`。10个accData对应10个人收集的加速度传感器数据，10个gyrData对应10个人收集的陀螺仪传感器数据。

 accData0.txt	2017/10/26 21:25	文本文档	1,336 KB
 accData1.txt	2017/10/26 21:25	文本文档	1,334 KB
 accData2.txt	2017/10/26 21:25	文本文档	1,333 KB
 accData3.txt	2017/10/26 21:25	文本文档	1,329 KB
 accData4.txt	2017/10/26 21:25	文本文档	1,341 KB
 accData5.txt	2017/10/26 21:25	文本文档	1,338 KB
 accData6.txt	2017/10/26 21:25	文本文档	1,335 KB
 accData7.txt	2017/10/26 21:25	文本文档	1,331 KB
 accData8.txt	2017/10/26 21:25	文本文档	1,329 KB
 accData9.txt	2017/10/26 21:25	文本文档	1,337 KB
 gyrData0.txt	2017/10/26 21:25	文本文档	1,394 KB
 gyrData1.txt	2017/10/26 21:25	文本文档	1,364 KB
 gyrData2.txt	2017/10/26 21:25	文本文档	1,402 KB
 gyrData3.txt	2017/10/26 21:25	文本文档	1,383 KB
 gyrData4.txt	2017/10/26 21:25	文本文档	1,378 KB
 gyrData5.txt	2017/10/26 21:25	文本文档	1,377 KB
 gyrData6.txt	2017/10/26 21:25	文本文档	1,398 KB
 gyrData7.txt	2017/10/26 21:25	文本文档	1,377 KB
 gyrData8.txt	2017/10/26 21:25	文本文档	1,409 KB
 gyrData9.txt	2017/10/26 21:25	文本文档	1,382 KB

 accData0.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
1508939178859 -3.7253778 1.3024458 10.429143
1508939178895 -3.6966474 0.82360536 10.515334
1508939178896 -3.169923 0.92895025 9.950302
1508939178907 -2.6719291 0.90979666 9.3661175
1508939178927 -2.6910827 0.48841715 9.11712
1508939178947 -2.873042 0.2777274 8.839393
1508939178967 -3.0550013 0.41180268 8.714894
1508939178990 -3.1028855 0.5363012 8.66701
1508939179007 -3.0933087 0.80445176 8.590396
1508939179027 -3.064578 1.2545617 8.6095495
1508939179047 -3.0262709 1.7429788 8.7532015
1508939179067 -2.930503 2.0590134 8.465898
1508939179088 -2.8921957 2.2409728 8.159439
1508939179108 -2.873042 2.2505496 7.97748
1508939179127 -2.7389667 2.1164744 8.226477
1508939179147 -2.6623523 1.8195933 8.456321
1508939179168 -2.7102363 1.7238252 8.475474
1508939179190 -2.7868507 1.8770541 8.379706
1508939179208 -2.8634653 2.030283 8.37013
1508939179227 -2.7485435 2.2026656 8.264784

对于每一个数据文件（加速度传感器数据和陀螺仪传感器数据），里面有30500行，即每一行对应一条数据。每一行有四个数字，以空格分隔，第一个是时间戳，第二个是X轴数据，第三个是Y轴数据，第四个是Z轴数据

请注意：虽然采集数据的时候设定了固定的采样频率50HZ，但是实际得到的数据并不是严格的间隔20ms。

你需要做什么

- ☐ 提供的数据是时间连续的序列，你需要自己将其分割成若干组数据，处理成为算法可以使用的格式用于训练和测试
- ☐ 标定数据。10个文件对应了10个人的数据，将处理完的数据的标签设置成为其对应的人
- ☐ 选择合适的机器学习算法，对处理好的数据进行分类（10分类任务）
- ☐ 不断调整算法的参数或结构进行优化以取得更好的结果
- ☐ 在实验的过程中随时记录你的实验情况，包括但不限于数据的预处理过程及结果、选择的算法、参数的设定、实验的训练过程和测试结果

实验要求

注意：为便于最终的效果对比，请务必遵守以下要求，否则将会影响成绩

1. 请自行将原始数据处理成为算法可用的数据并自行划分训练集、验证集和测试集。但是请保证 **测试集合** 中的数据量不少于**300组**
2. 划分训练集和测试集的时候请使用随机划分或者前N%为训练集，后(100-N)%为测试集的划分方式，不允许间隔划分或选取特殊位置的数据作为测试集
3. 请务必保证你提交的源代码 可以运行。因此请将处理数据的所有逻辑也一起提交
4. 对于提交的实验结果，请提供相关的运行 **截图** 或 **实验结果**

需要提交的材料

实验9-学号-姓名.[zip | 7z | tar]，其中包含：

1. 实验报告.[doc | docx | txt | pdf | md]，其中包括以下章节（不再使用之前的报告模板，括号内的百分比表示该部分内容的大概篇幅占比，你可以根据实际情况进行调整）：
 - 数据处理（30%）：请说明数据处理的方法，为什么要这样处理，并提供相关代码逻辑的截图
 - 实验过程（50%）：请说明实验的整个过程。主要包括 **算法** 的选择，算法相关参数或结构的设定，实验的 **思路** 和具体流程。
 - 结果及结论（10%）：展示实验结果。由于神经网络等算法每次运行的结果可能稍有不同，请提交效果最好的一次，但务必提交相关的截图或者运行结果。
 - 程序说明（10%）：请说明你的代码的运行方式和程序入口
2. **源代码**（文件夹），其中包括所有的源码。若有对原始数据的中间处理结果，也请放到这里

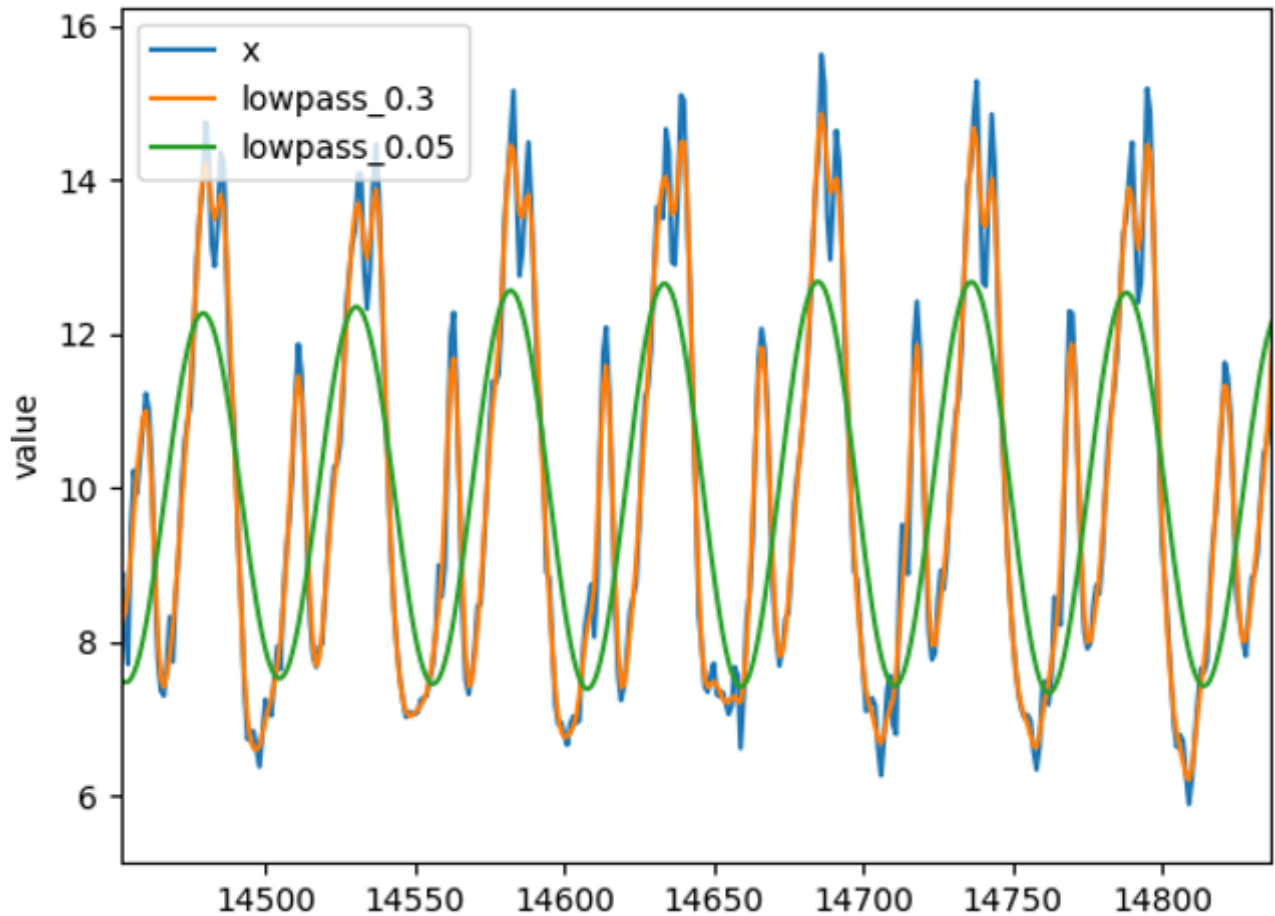
一些有用的提示

数据处理

数据预处理

在划分数据之前可以尝试对数据进行预处理，包括但不限于以下方法：

1. 滤波：对原始数据进行滤波可以有效地去除噪声平滑曲线 [2]。下图为使用巴特沃斯低通滤波器，截断频率分别0.3hz和0.05hz对原始传感器数据进行滤波



```
# 低通滤波代码
from scipy import signal
b, a = signal.butter(8, 0.02, 'lowpass')
filtData = signal.filtfilt(b, a, data) #data为要过滤的信号
```

2. 数据增强：可以使用插值等算法对数据进行增强处理。如使用RNN网络进行训练时，输入网络的每一个数据的数据点个数必须相等，但是划分数据之后每一个数据内的数据点个数可能不相等，此时可以通过插值处理得到固定数据点个数的数据

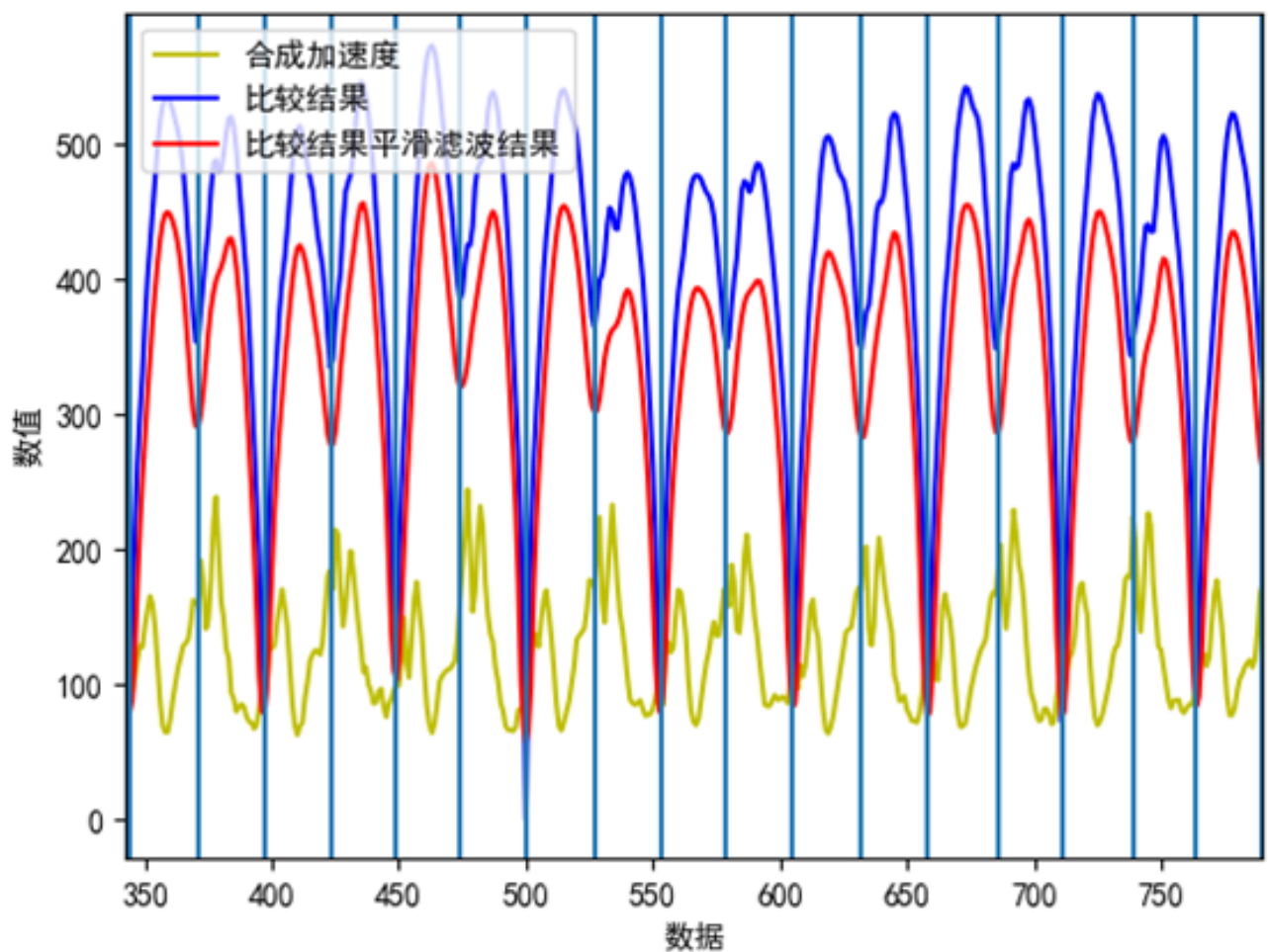
数据划分

需要将连续的时间序列数据划分为若干组数据，包括但不限于以下方法：

1. 滑动窗口分割：选择合适的窗口宽度和滑动步长，对数据进行划分

```
# 滑动窗口代码
with open(os.path.join(PATH/TO/DATA/FOLDER,"accData0.txt")) as acc_file:
    acc_datas = acc_file.readlines()
    for i,acc_data in enumerate(acc_datas):
        t,x,y,z = [float(i) for i in acc_data.split(" ")]
        acc_datas[i] = [x,y,z]
p = 0
while True:
    if p + WINDOW_LENGTH < len(acc_datas):
        yield acc_datas[p:p+WINDOW_LENGTH]
        p = p + WINDOW_LENGTH//2 # 每两个窗口有50%的重合部分
```

2. 步态周期分割：由于步态数据呈现明显的周期性，可以据此进行数据划分 [3]。如下图，选取一段数据作为模板，将模板在整个数据上滑动计算模板和数据之间的欧氏距离，得到呈现周期性规律的图像，据此划分步态周期



算法选择

原则上不对使用的算法进行限制。可以使用传统机器学习算法，如KNN，SVM等算法，也可以使用深度神经网络，对于网络的选择和结构也不做要求。给定在此数据集上进行分类一些常见算法的baseline结果，所以理论上你通过调整参数和网络结构所能取得的最好成绩应该比baseline要好。

算法	分类准确率(%)	说明
贝叶斯网络	90	
逻辑回归	92	
多层感知机	94	
KNN	94	K=1
KNN	93	K=2
KNN	94	K=3
J48	92	
随机森林	95	
SVM	92	
神经网络	97	1D卷积(5)+池化(3)+1D卷积(5)+池化(3)+全连接(10)
神经网络	95	GRU(128)+全连接(10)
神经网络	97	LSTM(32)+全连接(10)
神经网络	95	LSTM(32)+Dropout(0.01)+全连接(10)

如何获得更好的成绩

实验完成以下中的任意一条都会极大地提高你的成绩

- 获得最高的 分类准确率
- 对于数据中十个人里每一个，都提供了两个数据文件(`accDatai.txt` 和 `gyrDatai.txt`)，分别包含第 `i` 个人的加速度传感器数据和陀螺仪传感器数据。你可以使用单独一个文件中的数据进行分类。如果你同时使用两种数据，请 探索如何将两种数据有机地结合在一起使用，并探索相较于单独使用一种数据，同时使用两种数据是否会提升分类准确度，请通过详细的实验或者理论进行证明
- 通过调整网络结构或者改变模型参数，有效地提高分类准确率，并 从理论和实验两方面证明这种改进的意义

截止时间

2019-05-05 24:00之前将实验报告的压缩包文件提交至13061044@buaa.edu.cn

参考资料

- [1] 手机上到底有哪些传感器，它们都有什么作用呢？你知道吗？
- [2] 利用Python `scipy.signal.filtfilt()` 实现信号滤波
- [3] Muhammad Muaaz, Rene Mayrhofer: Smartphone-Based Gait Recognition: From Authentication to Imitation. IEEE Trans. Mob. Comput. 16(11): 3209-3221 (2017)