

机器学习工程师纳米学位

性别语音识别开题报告

赵普

2018年3月10日

项目背景

这个项目来自于Kaggle [Gender Recognition by Voice](#)，其目的是训练一个能判断一段音频信号是男性还是女性的模型。

性别语音识别属于[语言识别 \(ASR\)](#) 领域。语音识别技术所涉及的领域包括：信号处理、模式识别、概率论和信息论、发声机理和听觉机理、人工智能等等。研究该项目主要是自己对语音识别比较感兴趣，当下各种智能音箱也比较热门。

问题描述

这个项目要解决的问题是通过语音特征数据训练一个模型出来，使用该模型识别识别输入的声音特征并预测该声音是男性还是女性。

数据集和输入

项目所选用的数据集是Kaggle提供的数据，共包含3168个样本，其中男性和女性各占一半，均为1584个样本。数据集是原始数据经过处理后的音频特征，一共有20多个特征，并含有男性或者女性的标签。

数据集中有一些样本的某些特征为0，因此需要进行清洗；需要对数据进行处理并可视化，分析数据特点；为了进行测试和验证，还需对数据进行分割。

以下是特征和标签：

- meanfreq: 频率平均值
- sd: 频率标准差
- median: 频率中位数
- Q25: 频率第一四分位数
- Q75: 频率第三四分位数

- IQR: 频率四分位数间距
- skew: 频谱偏度
- kurt: 频谱峰度
- sp.ent: 频谱熵
- sfm: 频谱平坦度
- mode: 频率众数
- centroid: 频谱质心- peakf: 峰值频率
- meanfun: 平均基音频率
- minfun: 最小基音频率
- maxfun: 最大基音频率
- meandom: 平均主频
- mindom: 最小主频
- maxdom: 最大主频
- dfrange: 主频范围
- modindx: 累积相邻两帧绝对基频频差除以频率范围
- label: 男性或者女性

解决方法

项目推荐使用的模型：

- 逻辑回归
- 决策树
- 随机森林
- 支持向量机
- 神经网络
- XGBoost

基准模型

性别语音识别是一个二分类问题，该问题的Baseline（总是预测为男性）的准确率为50%。

评价指标

准确率：

$$accuracy = \frac{\text{正确分类的样本数}}{\text{总样本数}},$$

k折交叉验证准确率：

$$accuracy = \frac{1}{k} \sum_{i=1}^k \frac{\text{正确分类的样本数}}{\text{总样本数}},$$

项目设计

具体流程及步骤如下所示：

- 获取数据及导入数据（pandas）
- 分析数据特点，进行可视化
- 异常数据清洗
- 归一化、正态分布化
- 独热编码
- PCA主成分分析
- 训练模型（使用不同方法）
- 验证模型

引用

[1] <https://www.kaggle.com/primaryobjects/voicegender>

[2] <https://zh.wikipedia.org/wiki/语音识别>