语音性别识别

机器学习纳米学位开题报告

陈晓杰
2018年8月

1. 问题描述

该项目解决的是一个音频分类的问题，使用机器学习的方，判断一段音频信号是由男性还是女性发出的。该项目使用的数据中，每一条数据都由若干个音频特征和一个分类标识组成，所以本质上该项目解决的是一个监督学习[1]的分类问题。

2. 项目背景

说话人识别[2]是一项利用声音特征来识别人的语音识别技术。该项技术的历史可以追溯到四十年前，并在当时已经发现了不同个体间的语音特征的区别。在本项目中，需要解决的是根据语音特征来识别性别，是说话人识别的一种特例，因为最终的只需要分类成男性或者女性。

在机器学习的领域中，分类是它的一项主要的应用。在此项目中，数据集是已经从音频信号中提取出的特征和对应的分类标识组成的，因此可以利用监督学习的模型解决此问题，例如逻辑回归[3]、决策树[4]、随机森林[5]、SVM[6]、神经网络[7]、GBDT[8]和XGBoost[9]等算法。

3. 数据或输入

本项目中的数据集来自KORY BECKER在16年6月的语音性别识别项目[10]该数据集总共包含有3168个样本，其中50%为男性，50%为女性。样本数据是由音频文件解析的，这些音频文件来自男性和女性发言者。通过运用R语言的seewave和tuneR的包对语音样本进行了预处理[11]，分析频率范围为0hz-280hz（人类声音范围）。

4. 评估标准

判断一段语音是男性或是女性发出的，这是一个典型的二分类问题。在该项目中，适用的评估标准分类准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-score和样本预测时间。

- 混淆矩阵(confusion matrix)
  关于二分类问题中，测试数据集的真实值和与预测值，有以下四个关系：
  TN: True Negative，预测值为1，且预测对了

TP: True Positive，预测值为0，且预测对了
FN: False Negative，预测值为0，但预测错了
FP: False Positive，预测值为1，但预测错了

混淆矩阵的定义如下表所示：

	实际值为 1	**实际值为 0**
预测值为 1	TP	FP
预测值为 0	FN	TN

- 准确率(Accuracy)
  准确率可以直观地体现出分类器的性能，准确率越接近1，性能越好。根据混淆矩阵，准确率的定义是正确分类出来的样本数占样本总数目的比率：

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- 精确率(Precision)
  精确率也叫查准率，代表的是在所有被预测为1的样本中，真实值也为1的样本数目的比率：

$$Precision = \frac{TP}{TP + FP}$$

- 召回率(Recall)
  召回率也叫查全率，在医学上常常被称作敏感度，代表的是真实值为1的样本中，被正确预测数来的样本数的比率：

$$Recall = \frac{TP}{TP + FN}$$

- F1-score
  F1-score为精确率与召回率的调和平均值，越接近1，性能越好：

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

- 样本预测时间
  样本预测时间可以衡量分类器的对于数据预测的快慢，对于一个测试集，若其样本总数为N，分类器对测试集的预测总时长为T，则样本预测时间的定义如下：

$$t = \frac{T}{N}$$

5. 基准模型

为了判断项目中采用的机器学习的模型是否优于非ai领域的方法，可以用一个简单的算法作为基准模型[10]，这个算法不管输入音频的声音特征如何，将所有输入样本都分类为男性类别。

因为项目中的样本的标签中，男女各占50%，所以该基准模型的准确率将会是50%。更好的算法模型，准确率将会高于50%的这个基准值。

6. 项目设计

本项目采用随机森林的模型对数据集进行学习和分类，识别语音的性别。本项目的工作由以下几部分组成：

1. 数据准备：从kaggel[11]载数据集
2. 数据预处理：填补缺失值，分离特征和标签，将数据集划分为训练集和测试集
3. 创建分类器：采用sklearn的随机森林模型RandomForestClassifier创建一个分类器
4. 训练分类器：用训练集对分类器进行训练，由于随机森林算法的有放回的样本抽取方式会有OOB数据，可以不需要K-Fold CV，直接放入数据训练
5. 参数调整：用GridSearchCV对分类型的超参调整
6. 运用训练好的RF分类器对测试集的数据进行语音性别识别，计算其准确率和F1-score等性能指标，并对其预测效果进行分析。

---

1. https://zh.wikipedia.org/wiki/監督式學習 ↩
2. https://en.wikipedia.org/wiki/Speaker_recognition ↩
3. https://zh.wikipedia.org/wiki/邏輯迴歸 ↩
4. https://zh.wikipedia.org/wiki/决策树 ↩
5. https://zh.wikipedia.org/wiki/随机森林 ↩
6. https://zh.wikipedia.org/wiki/支持向量机 ↩
7. https://zh.wikipedia.org/wiki/人工神经网络 ↩
8. https://en.wikipedia.org/wiki/Gradient_boosting ↩
9. https://en.wikipedia.org/wiki/Xgboost ↩
10. http://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/ ↩
11. https://www.kaggle.com/primaryobjects/voicegender ↩