性别语音识别项目开题报告

领域背景

早在 50 年代就有涉及语音识别技术的研究，但那时还只是限制在约十个词汇量的单扬声器系统。随后断断续续的研究和发展，在 60 年代后期，苏联人发明了 Dynamic Time Warping (DTW) 算法，Hidden Markov Model (HMM) 算法也开始使用，并在 80 年代 HMM 替代 DTW 成为主流，当然还有其他一些算法，但都不够理想。直到 2010 左右，随着深度学习技术的兴起以及大数据和云计算的出现，将这些用于语音识别领域，克服了以前语音识别的诸多障碍，语音识别技术因此得以快速发展。如今，语音识别技术已比较成熟并有了许多应用，但语音识别是个很复杂的问题，依然有许多问题（如背景噪声、口音等）有待解决。

语音是人与机器最重要的交互方式之一。人们可以以语音交流的方式通过与手机上的语音助手（如 Google Assistant）来拨打电话、打开 app、访问网站等，也可以通过智能音箱（如亚马逊的 ecoh）播放音乐、控制智能家电设备、网上购物等，当然还可以通过其它的各种智能语音设备进行人机交互。而人与机器通过语音交互的核心点就是机器识别人类语言的问题（即语音识别问题），因此，在这方面的应用上，语音识别技术至关重要。

参考：https://en.wikipedia.org/wiki/Speech_recognition

问题陈述

在该项目中要解决的问题是通过说话者的一段语音判断该说话者的性别是男性还是女性。但计算机是难以直接通过音频信号来判断性别的，因此对于该问题又应分解成两部分，第一部分是将音频信号转化成多个可以用数值描述的特征，第二部分是利用这些特征来判断说话者的性别。对于第一部分，由于采用的数据集是已经用 R 语言从音频信号中提取特征后的数据集，因此不需要再做额外处理。此项目主要处理第二部分。通过特征判断性别，且性别只有男性和女性，所以这是一个可以用相应的机器学习算法来解决的二分类的问题。

数据集和输入

数据集 voice.csv (来源：https://www.kaggle.com/primaryobjects/voicegender) 包含 3168 个样本，其中 50%为男性，50%为女性。该数据集是已经使用 R 语言从音频信号中提取特征后的数据集。这些特征如下：

meanfreq	频率平均值 (in kHz)	centroid	频谱质心
sd	频率标准差	peakf	峰值频率
median	频率中位数 (in kHz)	meanfun	平均基音频率
Q25	频率第一四分位数 (in kHz)	minfun	最小基音频率
Q75	频率第三四分位数 (in kHz)	maxfun	最大基音频率
IQR	频率四分位数间距 (in kHz)	meandom	平均主频
skew	谱偏度频	mindom	最小主频
kurt	频谱峰度	maxdom	最大主频
sp.ent	频谱熵	dfrange	主频范围
sfm	频谱平坦度	modindx	累积相邻两帧绝对基频频差除以频率范围
mode	频率众数	label	男性或者女性

此数据集用来训练和测试机器学习模型。

另外，从实际生活中收集了 6 个 10s 的语音样本，并提前用 R 语言的 feature.R 脚本提取了特征，除了手动添加了 label 特征外，其它特征和 voice.csv 数据集的一样，保存在 real_voice.csv 中。此数据集用来看看训练好的模型对实际生活中的语音的预测效果怎样。

解决方案陈述

有带有标签的数据集，有目标分类，这是一个有监督学习的分类任务，只需选择一个有监督学习的分类模型训练数据并将训练好的模型来预测就解决问题了。在此项目中，我选择 XGBoost 模型。

基准模型

决策树是一个常用的、相对简单的、可解释性强的、效果也不赖的有监督学习模型，且也比较适合此项目问题，因此以我选择它做为基准模型。

评估指标

在此项目中，并不偏重某个分类，而是看整体的预测准确情况，因此，我选用准确率做为模型的评估指标。

项目设计

此项目此项主要按照如下步骤进行设计：

1. 数据探索：了解数据集的情况，包括有哪些特征，是否有缺失值，各特征值的类型是什么，若为数值型，那他们各自的常用统计量（如均值、中位数等）是多少，以及各特征之间的相关性如何等。

2. 数据准备： 根据数据集的情况及要使用的算法模型，做一些必要的数据前处理以便能更好的训练模型。这是个二分类问题，可以把目标分类转化 0 和 1，以方便处理。由于我采用的模型的基学习器是决策树，而决策树对大数值范围、异常值有很好的鲁棒性，因此对数据集无需做标准化处理或异常值处理。

3. 基准模型训练：训练出一个基准模型做为基准，以便对解决方案模型的好坏情况有一个参考。

4. 模型训练：设置模型超参，对模型进行初步训练。

5. 模型评估：评估模型的表现情况。

6. 模型优化：调节模型超参，对模型进行优化。打算使用网格搜索法对模型调优。

7. 模型测试：对最终模型进行评估测试。

8. 应用测试：拿最终模型对从实际生活中收集的语音进行预测，看其在实际应用中的表现如何。