

机器学习（进阶）工程师纳米学位毕业项目

语音性别识别

Max Liu 刘子铭*

2018年8月24日

*电子邮件:simonliu245@gmail.com

目录

1	问题定义	3
1.1	项目概览	3
1.2	问题说明	3
1.3	评价指标	3
2	问题分析	4
2.1	数据研究	4
2.2	探索性可视化	5
2.3	算法与方法	6
2.3.1	单个监督模型简介	6
2.3.2	融合模型	8
2.3.3	技术	8
2.4	基准测试	8
3	方法	9
3.1	数据预处理	9
3.2	算法实施	10
3.2.1	单个监督模型训练	10
3.2.2	融合模型训练	12
3.3	改进	12
4	结果	13
4.1	模型评估与验证	13
4.2	结果分析	14
5	结论	14
5.1	预测失败样本可视化	14
5.2	思考	15
5.3	改进	15
6	参考文献	16

1 问题定义

1.1 项目概览

该项目来自于Kaggle¹。简单的说，项目的目的很简单，就是去判断一段音频信号属于男性还是女性，这是一个监督学习中的二分类问题。

从大的领域上看，语音的性别识别还仅仅是语音识别领域的一个小分支，且技术上已经成熟，人们已经开始关注更精细的问题²。自20世纪70年代人们创立隐马尔可夫（HMM）模型以来，语音识别领域就开始逐步发展。其核心技术包括隐马尔可夫模型（HMM），混合高斯模型（GMM），MFCC，LM等等。现如今，在语音识别领域效果最好的当属深度学习技术³，其错误率在以往最好的系统GMM-HMM框架的基础上相对还要下降30%或更多，这一成就也使得语音识别技术有了更强的实用性。

从小的方向上看，实际上性别的识别是语音识别领域的一个小方向，属于语音识别中单个特征的识别问题³。虽然它的功能很单一，但是性别识别的应用却不简单。毕竟，一旦事先知道了男女性别，很多问题就方便得多了⁴。在如今Siri，Cortana等语音交互式应用炙手可热的趋势下，语音识别仍保持相当的重要性。因此，语音性别识别的确仍然是一个值得研究的课题。

我在该项目中使用的是多个监督模型以及一个它们的简单融合模型。由于时间有限，我并没有使用Stacking或者Blending融合模型，而是使用投票的机制⁵。在简单模型的优化上，我试图通过单个监督模型的优化来提高整体性能。

最后提一下项目使用的音频信号，这是已经经过seewave和tuneR软件包处理过的¹。数据集包含3168个样本，收集自男性和女性发言人，分析频率范围为0hz-280hz（人类声音范围），其中50%为男性，50%为女性。

1.2 问题说明

该语音性别识别项目的基本任务是准确判断一段音频信号属于男性还是女性。

之前已经提过，这是一个二分类问题，是监督学习（supervised learning）的问题，学习的目标是音频信号对应的性别标签。

可以想到的是，如何提高判断的准确性是一个重要问题，这也正是我们看重的评价指标。

1.3 评价指标

我们使用的评估指标是准确性（Accuracy），与Kaggle¹上的相同，其计算公式如下所示：

$$\begin{aligned} \text{Accuracy} &= \frac{\#True\ Positive + \#True\ Negative}{\#True\ Positive + \#True\ Negative + \#False\ Positive + \#False\ Negative} \\ &= \frac{\#Rightly\ Predictions}{\#All\ Samples} \end{aligned}$$

不过需要注意的是，在真正操作的时候，我们关注的是测试集上的准确率而非整个数据集。因此，我会给出模型在不同数据子集上的评价。

当然，除此之外，我们还有诸如Recall，Precision，F1 measure这样的参数⁶，这些参数将

会在单个模型的训练中给出，利用训练所得分类器的方法可以方便地返回这些参数。它们将用于更客观地判断模型表现。

2 问题分析

2.1 数据研究

首先，该项目用到的数据集来自Kaggle，数据集包含3168个样本，男性样本和女性样本都存在，简单测试可知，的确，男女样本各占50%，分别有1584条。

下面给出数据集中包含的特征（共20个）：这些特征是已经经过R脚本提取出的声学特

表 1: 数据集特征及其含义

meanfreq	频率平均值(in kHz)
sd	频率标准差
median	频率中位数(in kHz)
Q25	频率第一四分位数(in kHz)
Q75	频率第三四分位数(in kHz)
IQR	频率四分位数间距(in kHz)
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

征，用于label标签的判断。可以观察到的是数据集的特征部分的值均为数字，只有标签（性别）部分的值为male或female，这不利于后续处理，因此我们必然要进行独热编码。

另外从数据集的统计特征上看，譬如方差和最大值最小值，数据集对于特征kurt存在明显极值，数据波动范围较大的特征是kurt。skew、maxdom和drange虽然也有小幅波动，但波动范围比kurt都小得多。不过这些不应该算是极端值，因为根据这些特征的统计意义（表1），这些音频信号即使波动比较大，也都是可以发生的。另外，在数据集的预览中可知并无缺省值。

2.2 探索性可视化

在分析问题之前我们可以判断，在声学的意义上，一定有一些特征是可以区分男女声音的，否则该问题也就失去了意义。

从生活经验上判断，男性的声音普遍雄厚低沉，而女性声音尖细高亮⁷。因此，根据上述特征的描述，它们多为频率相关的特征，是很可能将声音成功分类的。为此，我们可以筛选重要特征并将性别作为标签进行可视化。这里的代码来自数据预处理部分，经过筛选，我们的确得出了重要特征并作了可视化。其中重要的特征是平均基音频率和频率四分位数间距，细节请见数据预处理部分。

首先，我们做了如下的特征可视化。由于这里的数据集含有3168条数据且共有20个特征，若使用`pdf`格式绘图，我的设备就会卡住，因此在这里我先不使用`pdf`模块，只导出`png`格式的图片：

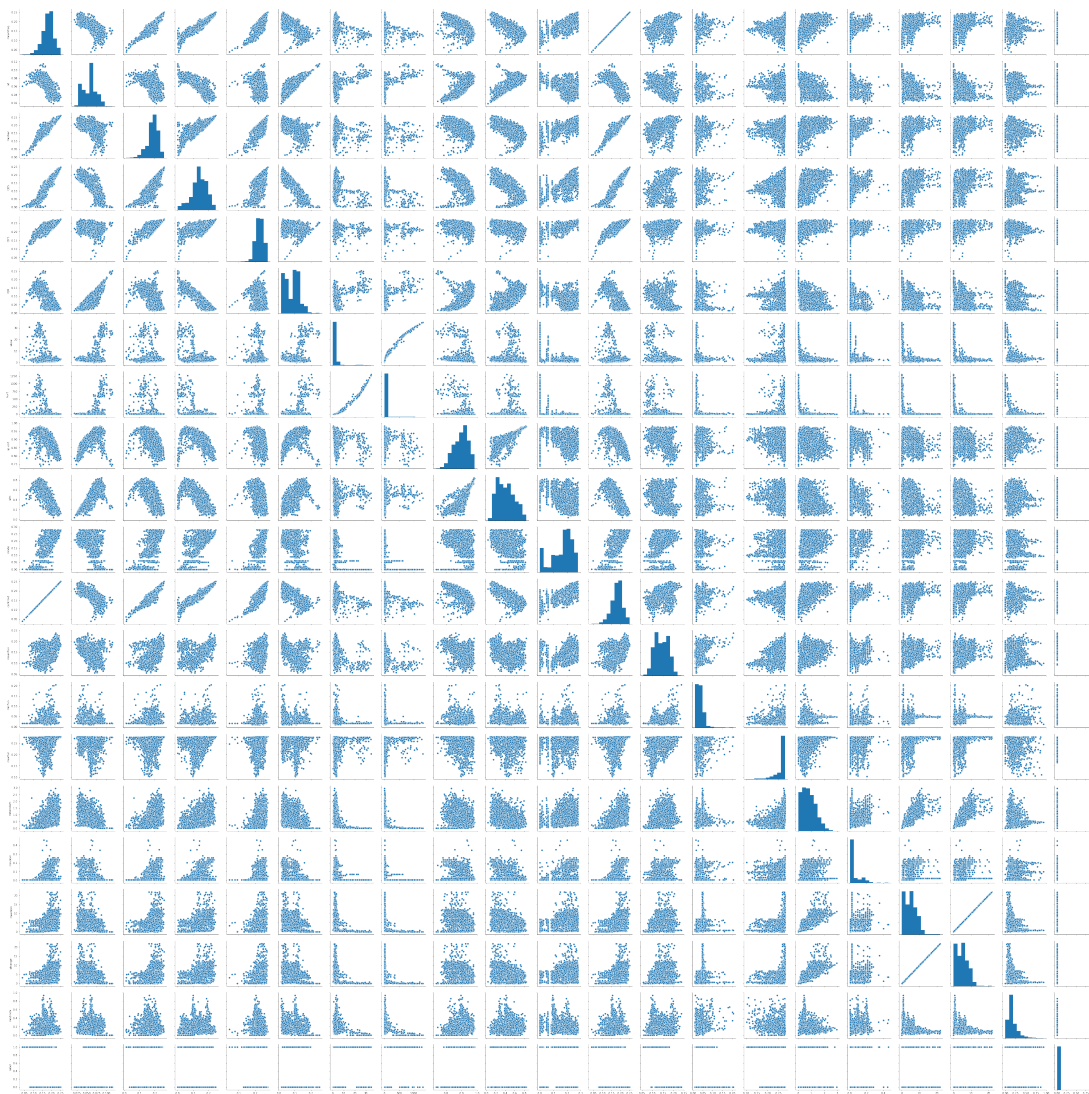


图 1: The Distribution of Features and Gender

这是数据集的分布情况，包括特征和性别标签，其中对角线上的图就是自分布，其

它位置的图为不同特征或标签之间的分布。可以看到，有些特征几乎取值集合很小，比如skew和kurt，这样的特征对于分类的效果可能很弱。其它的特征分布则较为广泛，各个区间上都有一定的样本数量。

下面则是经过计算相关系数后得出的两个重要特征的可视化结果，具体步骤见数据预处理部分。

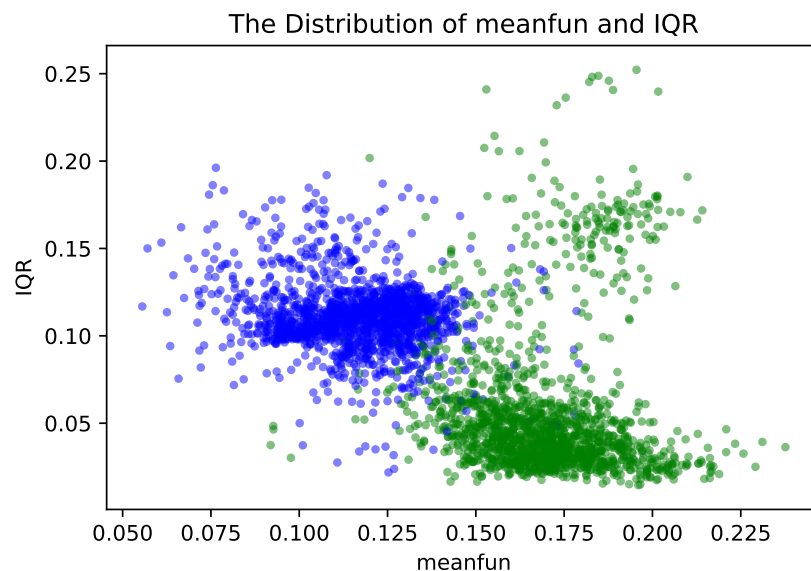


图 2: The Distribution of meanfun and IQR

显然，这两个特征可以较好地将数据点分为两类，其中蓝色点代表男性样本，绿色点代表女性样本。

2.3 算法与方法

2.3.1 单个监督模型简介

我们使用的单个监督模型一共有8种，都可以执行二分类的任务。下面依次简介这些模型的原理⁸。

- 基准模型（Baseline Model）

基准模型的使用是相当于设置了一个阈值，可以判断其它已经训练好的模型是否能够得到比基准模型更好的效果。

这里我们使用的基准模型是一个“天真”的预测器，无论输入的语音信号特征为何，它都会给出“男性”的预测。

显然这个“天真”预测器是合理的，因为按概率来说，输入的样本的实际标签为男性或女性的概率都是50%，所以基准模型的准确率应该是50%，这与人类随机猜测的结果相同。因此，“天真”预测器作为基准模型是合理的，如果训练的模型的准确性还不如它，那这个模型就是失败的。

- Logistic回归（分类）模型

Logistic回归本质上是一种线性分类器，其基本假设是特征和分类结果存在线性关系。由于在二分类的时候要将结果映射到0 或1。因此，在将特征向量线性处理后还需要一个映射把结果映射到(0,1)区间上。使用Logistic函数就是Logistic 回归（分类）模型种Logistic的由来。

其参数的学习方式有两种，一种是精确计算的解析方法，原理参见Stand Ford的课件¹³；另一种是随机梯度上升，即下面的SGD分类模型。

- 随机梯度参数估计（SGD）模型

这种模型其实就是Logistic回归模型的一种，但是计算的精度可能并不是那么高，原理已经在上一条中描述。

- K近邻（分类）模型

K近邻本身十分简单，容易实现，用通俗的话说就是“近朱者赤，近墨者黑”。由于K近邻算法受异常值干扰的影响比较大，因此它的效果可能不是太好。在处理的过程中，我会根据它的表现决定最后模型融合的时候是否采用K近邻。

- 支持向量机（SVC）模型

支持向量机的核心就是最大限度地不同类别分开。它的原理决定了它的优点：泛化能力强，且可以有效地处理高维数据，而这些优点很适合该项目（20个特征构成的特征空间）。

只是由于SVC对参数调节和核函数的选择敏感，受噪声的影响也较大，所以处理二分类问题时要调整它的参数，这会消耗较多的时间。

支持向量机的核函数是它处理高维数据的能力源泉。一旦数据变得复杂，问题很可能成为线性不可分的情况。支持向量机会先在低维空间中计算，然后核函数会把低维数据映射到高维特征空间，在高维特征空间问题就是线性可分的，从而可以得到线性超平面。常见的核函数包括多项式核、高斯核和线性核，它们都可以作为合适的高维映射¹⁴。

- 分类和回归树模型（CART）

对于这个项目，决策树也是可行的。决策树有着诸多优点，本身简单，对数据的处理不复杂，便于实现，可以较快地处理大量数据。同时，决策树的计算复杂度较低，而且关于缺失数据的影响较小。这决定了它在该项目上的优势，在后面的结果中可以看到，决策树的表现是比较优秀的。

- 随机森林模型

随机森林实际上就是决策树的集成，是一种集成算法，也可以把它看作是决策树的模型融合。随机森林的每个基本单元，即一颗决策树，都会对输入数据产生一个预测结果，多颗决策树通过投票方式决定投票次数最多的输出为最终结果。综合森林的输出，它的效果一般比决策树本身更好。

- 多层感知分类模型

这是一个神经网络模型，即多层神经元组合成的预测模型。每一层神经元的输出都会作为下一个神经元的输入，通过是否调整自身权重或是每个神经元的激活函数，它最终会返回一个输出，这就是该模型的直观原理。一般来说，对于数据量很大的任务，它的学习能力会更强。

- XGBoost模型

XGBoost和随机森林一样，都以决策树为基础。二者区分的地方在于XGBoost在损失函数中额外添加了正则项，可以控制模型的复杂度，以及使用的树的集成方式与随机森林略有不同。

除此之外，XGBoost优于传统的梯度提升决策树算法，毕竟，它对损失函数做了二阶Taylor展开而梯度提升决策树只使用了一阶导数。XGBoost有诸多优点使得它能够在最近的一些项目中取得更优异的成绩¹⁵。

2.3.2 融合模型

由于时间的关系，这里只使用最简单的投票（Voting）方法来实现模型的融合⁹。

首先在单个监督模型训练结束后，我会选出效果最好的几个模型。对这些模型再进行投票选择就得到最终的融合模型了。

2.3.3 技术

本项目使用的技术来自于经典的Python编程库，包括Scikit-learn、pandas⁸、Numpy、Scipy、Matplotlib和seaborn。项目还基于Anaconda提供的Python2.7编程环境，用Jupyter Notebook的形式展示代码实现的结果。这些都是一直在维护的非常成熟的机器学习编程库，对项目的实现提供了便利的环境。

由于模型众多，单个模型的参考文档就不在该报告中给出了，而是在Jupyter Notebook生成的ipynb文件中提供了链接。如果有重要的技术并非来自于上述机器学习编程库，我将给出标记并在参考文献中给出链接。

2.4 基准测试

该项目已经提供了基准模型，但是50%的准确率显然不是我们想要的，因此这里参考了Kaggle上已有的模型训练的结果：

表 2: Accuracy on Kaggle

Baseline (always predict male)	50% / 50%
Logistic Regression	97% / 98%
CART	96% / 97%
Random Forest	100% / 98%
SVM	100% / 99%
XGBoost	100% / 99%

这里我并没有试图精细优化单个模型，而是想看看简单的投票融合模型能够达到怎样的效果，初步的要求是性能要优于单个的基本监督模型，即准确率能够达到96%以上。这样的效果就已经可以解决绝大多数的问题了，至于未能预测成功的样本，我也会给出解释和改进的方法。

3 方法

3.1 数据预处理

首先我们导入csv格式的数据集并给出数据集的一些基本属性和个别条数据的预览。显然是数据集包括了3168条数据，分为20个特征和1个标签，即我们将要预测的性别。

在数据的预览中我们可以得到以下3条结论，这也对应了3步处理措施：

- 我们的预测目标label是字符格式，为了方便处理，我们需要独热编码；
- 数据集的前5组数据的label均为male，这说明数据可能排列并非随机，因此我们可能需要想办法打乱数据顺序再进行分割，用Excel软件预览csv文件，发现数据集的前半部分均为male，后半部分则为female，这证实了该想法的正确性；
- 第一条数据的倒数后两个特征dfrange和modindx均为0，这可能是数据缺失所致，我们也许需要先剔除缺失的数据，以免对预测器有所影响。不过，我们后面会分析如何排除此处数据缺失的可能性。

完成对性别标签的独热编码（male记为1，female记为0）后我思考了是否存在缺失数据，尽管根据数据集提供者的描述，数据集是规整的。

简单地用Excel工具查找了数据集中值为0的数据发现，值为0的数据共有366个（指的是366个特征数据），它们既可能是被0填充后的缺失数据，也可能是参数本身就为0。因此，是否处理它们还有待分析。进一步观察表明，这些值为0的数据有以下两个分布特征：

- 它们多出现在mode，dfrange和modindx列；
- 它们多成片出现，常常有几条排列在一起的数据出现大量0。

不过，根据Kaggle上的参数解释也就释然了：

表 3: 被怀疑存在缺失的特征意义

参数	参数解释
mode	频率众数
dfrange	主频范围
modindx	累积相邻两帧绝对基频频差除以频率范围

频率众数在音频信号大多数时间内没人说话的时候可以0；主频范围为0或者绝对基频频差也是可能为0的。所以从其声学意义看，这些数据为0是比较合理的，毕竟366个特征数相对于整个数据集还是比较少的。据此，我认为还是没有必要把它们认定是缺失数据。

接下来我还做了数据顺序的打乱，以免将来分割数据集的时候出现某个性别的数据出现占比过大的情况。

在监督算法实现之前我还做了一些简单的数据处理，包括重要特征的查找和标准化处理。经由pandas中的corr()函数做特征和性别标签相关性的分析，我得到了与性别标签相关性最强的几个特征。与性别标签相关性最强的是meanfun，相关系数是-0.833921；第二强的是IQR，相关系数是0.618916；其余特征的相关系数大于0.5的只有Q25。因此，我们可以选择前两个属性进行可视化操作，即之前探索性可视化中的结果。这里就不做展示了

从第6页的图1可以观察到，这两个特征的确很好地划分了数据集，蓝色点为男性语音样本，绿色为女性样本，呈现出了明显的二分类现象。这进一步证实了问题的可解性。

当然，数据集的标准化也很重要，这样可以消除量纲对目标的影响。经标准化处理后，不同量纲的特征会有相同的尺度（scale）¹⁰。这样，在学习参数的时候，不同的特征对参数的影响程度才会一样，参数也会尽快收敛。

3.2 算法实施

这一部分主要是训练常见的监督模型，训练的思路参考了一个使用R语言分析数据集的例子¹¹。

3.2.1 单个监督模型训练

下面列出我所使用的模型：

- 基准模型
- Logistic回归（分类）模型
- 随机梯度参数估计（SGD）模型
- K近邻（分类）模型
- 支持向量机（SVC）模型
- 分类和回归树模型（CART）
- 随机森林模型
- 多层感知分类模型
- XGBoost模型
- 融合模型（Vote型）

当然，在实现算法之前需要对数据集做训练集、验证集和测试集划分，这是监督学习的惯用方法了。这里每次切分的比例都是5:1，即先随机分割80%做训练集，20%做测试集；再从原始训练集中分出80%做最终训练集，其余20%做验证集。也就是说，64%的数据集为训练集，

包括2027个样本；16%的数据集为验证集，包括507个样本；20%的数据集为测试集，包括634个样本。

训练的结果在验证集上做了初步测试，对比Kaggle上的最优结果如下。其中Kaggle上的最优准确率结果为训练集/测试集，这里的结果为训练集/验证集。测试集在最后使用。

表 4: 准确率初步对比

模型	Kaggle	Model Trained
Baseline Model	50% / 50%	50.0% / 49.9%
LR	97% / 98%	97.4% / 97.6%
SGD	Not Trained	97.0% / 97.0%
KNC	Not Trained	98.7% / 96.8%
SVC	100% / 99%	98.5% / 98.2%
CART	96% / 97%	100% / 97.2%
Random Forest	100% / 98%	99.9% / 99.0%
Neural Network	Not Trained	98.7% / 97.2%
XGBoost	100% / 99%	99.7% / 98.6%

从以上结果可以看到，单个监督模型的效果还是很好的，它们在验证集上的准确率普遍要比训练集上的高。而且从Notebook中其它参数的情况看，单个模型的可信度是很高的，只有一小部分的预测会出现问题。

这里我还要插上一个分析，有关我在开题报告和数据预处理中查找重要特征部分所做过的假设，即借用sklearn中随机森林分类器的feature_importances_函数来得到特征重要性。下面就是重要性的可视化结果。

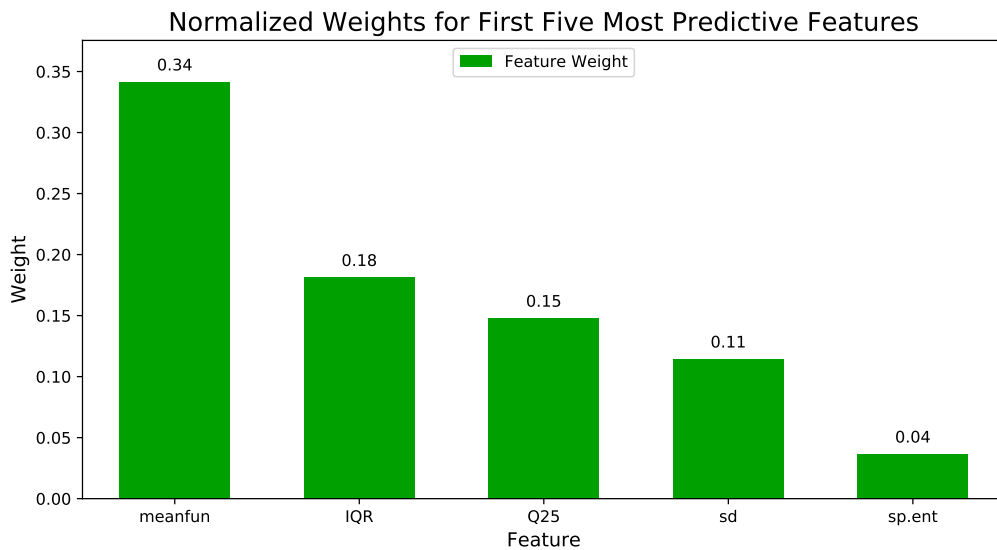


图 3: Normalized Weights for First Five Most Predictive Features

从此结果上看，meanfun特征和IQR特征的确是最重要的两个特征，与相关性分析的结果相

同。二者分别占比34%和18%，总和超过了50%，这进一步验证了特征重要性的结果。

同时，XGBoost也提供了类似的函数`plot_importance()`，其特征重要性的结果如下：

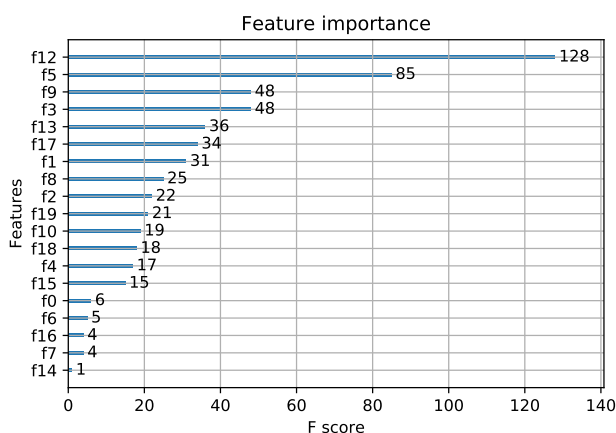


图 4: Importance of Features by XGBoost

图中f12指数据处理时特征列表中第12个特征，即`meanfun`，同理f5指的是`IQR`，这进一步强化了结果。实际上XGBoost和随机森林确定特征重要性的方法是不同的，随机森林每次抽样构造决策树都留下一些未选则的样本，称为袋外数据，随机森林选择用袋外样本的误差来评估特征的重要性；而XGBoost简单地统计不同特征在所有树中出现的次数（即节点划分的次数），把该次数作为重要性的评分。

3.2.2 融合模型训练

这里我只使用了最简单的投票型融合模型。其原理是使用多个训练好的分类器，对每一次测试，它们各自都进行一次预测，预测结果最多的一个当选为最终结果。这样做的优点很显然，可以避免某一个模型的偶然错误，各方投票决定的结果往往准确性更高。

根据上述训练结果在验证集上的表现选出最优的5个分类器，分别是SVC、CART、随机森林、多层感知分类模型和XGBoost模型（排名未分先后）。得到投票函数后仍用验证集进行测试。

3.3 改进

承接上文，新的融合模型在训练集上的表现仍然很好，准确性为99.7%，在验证集上的表现仅次于随机森林，与XGBoost持平，准确性为98.6%。

但是这似乎还不够，我们仍然有其它方法可以得到更好的结果。我将提出两种方法并实现其中较简单的一种。

第一种方法是使用更强的融合模型，如Blending或Stacking，由于时间限制，我不打算实现这一部分，只把方法留在这里。另一种就是我要实现的，通过改进单个预测器来改进融合模型。改进的方法是网格搜索和交叉验证法，希望找到单个模型的更佳参数，通过优化后的单个模型改进融合的投票模型。

通过查阅sklearn文档可以确定待优化的参数及范围，经过网格搜索和交叉验证，我们得到了新的单个监督模型，再做同样的投票融合，即可得到优化后的融合模型。

4 结果

4.1 模型评估与验证

在测试集上测试后，我们列出到现在为止所有模型的准确率：

表 5: 准确率初步对比

模型	Kaggle	Model Trained(Training, Validation, Testing)
Baseline Model	50% / 50%	50.0% / 49.9% / 50.0%
LR	97% / 98%	97.4% / 97.6% / 97.1%
SGD	Not Trained	97.0% / 97.0% / 96.1%
KNC	Not Trained	98.7% / 96.8% / 97.9%
SVC	100% / 99%	98.5% / 98.2% / 98.0%
CART	96% / 97%	100% / 97.2% / 97.2%
Random Forest	100% / 98%	99.9% / 99.0% / 97.0%
Neural Network	Not Trained	98.7% / 97.2% / 97.9%
XGBoost	100% / 99%	99.7% / 98.6% / 97.2%
Voted Classifiers	Not Trained	99.7% / 98.6% / 97.6%
CVed Voted Classifiers	Not Trained	99.1% / 97.8% / 97.6%

另外，我给出不同模型训练所消耗的时间：

表 6: 模型训练消耗时间

Models	Time cost(second)(Primary Model Training/ Grid Searching)
Baseline Model	None / None
LR	0.0280001163483 / None
SGD	0.00200009346008 / None
KNC	0.00300002098083 / None
SVC	0.197999954224 / 242.392999887
CART	0.0250000953674 / 1.04200005531
Random Forest	0.0450000762939 / 47.8340001106
Neural Network	0.87299990654 / 1114.97600007
XGBoost	0.275000095367 / 181.176000118

从上表可以看到，网格搜索和交叉验证所消耗时间普遍远超原始模型训练。我还会在README文档中重述此时间表。

4.2 结果分析

从以上结果来看，可以看到经过初步优化，经过交叉验证和网格搜索后的融合模型在测试集上保持了之前同样的准确率97.6%，这是为什么呢？

据我分析，输出新的单个分类器可以知道，交叉验证的确调整了预测器的参数，模型的确是做了一些优化；但是我们的测试集的规模比较小，预测失败的数据可能是比较顽固的数据，即可能处在男女声音的边界处，呈现出的是不易判断男女性别的语音。因此即使模型的性能有所提升，它在顽固数据上的表现也不一定好。这些点会是最后改进的关键之一。换句话说，这里的融合模型没有带来提升，是因为模型本身存在一定的不稳定性，尤其是由于这里的数据集较小，而且该任务的预测精度需求高，所以某种程度上说会有一定的偏差是无法避免的。

总之，这就是上述5个模型利用投票机制融合得到的最优模型。它在测试集上的准确率为97.6%。我想，如果训练集的数据充足，且测试集中处于男女语音边界处的数据较少的话是可以得到更高的准确率的。

虽然模型在测试集上的表现并未达到Kaggle上最优的结果，但是97.6%的准确率已经足以解决绝大多数的语音性别识别问题了，预测会失败的数据恐怕就是那些人类也很难分辨的语音了。因此，我认为，模型已经足够优秀。

5 结论

通过以上的模型训练，可以确认的是该项目的基本任务已经完成。最终得到的预测器在测试集上达到了97.6%的准确性。这与Kaggle上的最优结果相比还有差距，但这已经达到了较好的预测表现。鉴于这里测试集较小，我认为该模型的性能在应用上会更强。

5.1 预测失败样本可视化

这一部分是对于预测失败数据的讨论，我们把测试集的meanfun和IQR参数作为横纵轴的坐标，把性别投射到二维平面上，结果如下：

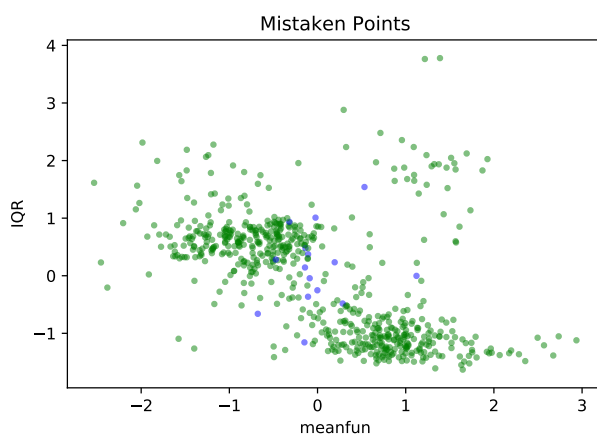


图 5: Mistaken Points

如上所示，可以看到少量蓝色的数据点是预测失败的数据。它们普遍分布在男性、女性样

本点集合的交界处，预测失败也是可以理解的。毕竟现实生活中很多男女的声音也是难以分辨的，因此总的来说我们的模型的结果已经很好了。这些难以分辨的点可能是算法所需要克服的。

5.2 思考

根据上述预测失败数据的可视化，我想这些难以分辨的点可能是算法需要克服的，而我们的一般监督模型已经达不到这样的要求了。所以我所思考的是如何判断这些语音数据的性别归属。

我的想法有两点。第一是通过深度学习强大的学习功能抽象出男女声音的标志，但是这要注意训练的时间控制以及拟合的效果；第二是我在开题报告中提到的SMuRFS算法¹²，由它选择出与性别密切相关的声学特征并重新建模，往往这会得到多维重要特征，在新的维度上，这些难以分辨的数据可能被成功识别。

5.3 改进

改进的主要方向是准确分辨难以预测成功的语音信号，这些信号的混淆能力很强，有时候人类也无法分辨。

后续对模型的改进措施还是很多的，在模型方面，我们可以尝试其它boosting类模型如lightgbm；在特征层面，可以试图交叉特征以构造新的特征，它们可能会与性别标签有更强的相关性。当然，我在思考中提出的也是改进措施。

6 参考文献

- [1] Kory, Becker. Gender Recognition by Voice—Kaggle[EB/OL].
<https://www.kaggle.com/primaryobjects/voicegender>.
- [2] 腾讯AI Lab副主任俞栋: 语音识别领域四大前沿问题亟待研究[EB/OL].
<http://tech.qq.com/a/20170528/017297.htm>.
- [3] 俞栋, 邓力. 解析深度学习-语音识别实践[M]. 北京:电子工业出版社, 2016.
- [4] 台达电子. 语音技术_性别辨识和语者验证_语音识别_CTI 论坛[EB/OL].
http://www.ctiforum.com/factory/list/www.delta.com/delta10_1105.htm.
- [5] 立刻有. Ensemble Learning-模型融合-Python实现- CSDN博客[EB/OL].
<https://blog.csdn.net/shine19930820/article/details/75209021>.
- [6] 范淼, 李超. Python机器学习及实践——从零开始通往Kaggle竞赛之路[M]. 北京:清华大学出版社, 2016. 41-43
- [7] 比的原理. 解读男女声音的区别:亮度,糙度— 比的原理[EB/OL].
<https://bideyuanli.com/p/863>.
- [8] 范淼, 李超. Python机器学习及实践——从零开始通往Kaggle竞赛之路[M]. 北京:清华大学出版社, 2016.
- [9] Triskelion. Kaggle Ensembling Guide — MLWave[EB/OL]. <https://mlwave.com/kaggle-ensembling-guide/>.
- [10] 数据什么时候需要做中心化和标准化处理? - 知乎[EB/OL].
<https://www.zhihu.com/question/37069477>.
- [11] Kory, Becker. Identifying the Gender of a Voice using Machine Learning[EB/OL].
<http://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/>.
- [12] Mayer J, Rahman R, Ghosh S, et al. Sequential Feature Selection and Inference using Multivariate Random Forests[J]. Bioinformatics, 2017, 34(8).
- [13] Andrew, Ng. CS229 Lecture notes[EB/OL].
<http://cs229.stanford.edu/notes/cs229-notes1.pdf>.
- [14] July. 支持向量机通俗导论理解SVM的三层境界[EB/OL].
https://blog.csdn.net/v_july_v/article/details/7624837.
- [15] 机器学习算法中GBDT 和XGBOOST 的区别有哪些? - 知乎[EB/OL].
<https://www.zhihu.com/question/41354392/answer/98658997>.