Machine Learning Nano Degree
Gender Recognition of Voice

Jiacheng Shen

March 15, 2018

Contents

1 Definition

1.1 Project Overview

Voice recognition has a long history. In 1952, Bell Lab designed a system to recongize single digit. Then, great progress was made by methods like Linear Predictive Coding, Dynamic Time Warp and Hidden Markov Model. Nowadays, Machine Learning models like RNN are applied.[3] For example, the famous Chinese company iFLYTEK can reach an accuracy of 98%.[1]

In this project, the goal is to build a model to recongize the gender of voice. The dataset can be found on Kaggle[2]. The raw wave files have been extracted features by R. There are 1584 female samples and 1584 male samples. Finally, a web service will be constructed based on the trained model as an application.

1.2 Problem Statement

This project is a **suprivised binary classification** problem. The voice can be classfied to female or male. The goal is to use the given features of voice, build a model and recongize the gender of voice.

1.3 Merrics

Since the dataset is balanced (half female and half male) and the importance of both genders are the same, so recall and precision rate mean nothing. So accuracy is chosed to evauate the model. What's more, training and predicting time is also considered because the model need to be deployed on CPU server.

1. **Accuracy**. Accuracy counts how many samples are predicted correctly.

$$\text{Accuracy} = \frac{\sum \text{Correctly predicted}}{\sum \text{All Samples}} \times 100\%$$

2. **Time**. To construct a web service on a normal CPU server, training and predicting time is also taken into account. The user cannot wait too long for the response.

2 Analysis

2.1 Data Exploration

The dataset is downloaded from Kaggle. There are 22 features and 3168 samples. It is a balanced dataset with 1584 females and 1584 males.

As shown in Table 1, there are 22 features about the voice and the last one is the label, which need to be predicted female or male.

More details about the features are shown below.[2]

1. meanfreq: mean frequency (in kHz)
2. sd: standard deviation of frequency
3. median: median frequency (in kHz)

Category	Names						
Frequency	meanfreq	sd	median	Q25	Q75	IQR	
Spectrum	skew	kurt	sp.ent	sfm	mode	centroid	peakf
Fundamental Frequency	meanfun	minfun	maxfun				
Domain Frequency	meandom	mindom	maxdom				
Range	dfrange	modindx					
output	label						

Table 1: features

4. Q25: first quantile (in kHz)
5. Q75: third quantile (in kHz)
6. IQR: interquantile range (in kHz)
7. skew: skewness (see note in specprop description)
8. kurt: kurtosis (see note in specprop description)
9. sp.ent: spectral entropy
10. sfm: spectral flatness
11. mode: mode frequency
12. centroid: frequency centroid (see specprop)
13. peakf: peak frequency (frequency with highest energy)
14. meanfun: average of fundamental frequency measured across acoustic signal
15. minfun: minimum fundamental frequency measured across acoustic signal
16. maxfun: maximum fundamental frequency measured across acoustic signal
17. meandom: average of dominant frequency measured across acoustic signal
18. mindom: minimum of dominant frequency measured across acoustic signal
19. maxdom: maximum of dominant frequency measured across acoustic signal
20. dfrange: range of dominant frequency measured across acoustic signal
21. modindx: modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range
22. label: male or female

The Table 2 showes the head and the tail of the dataset. The features except label is all continuous number. The first half are all males and the second half are all females. Therefore, the dataset need to be randomly rearranged to break this distribution.

According to the Table 3, the distributions and ranges of features differ a lot from each other. Thus, normalization is needed to prevent preferrence on large features.

Features	Sample 0	Sample 1	Sample 3166	Sample 3167
meanfreq	0.059781	0.0660087	0.143659	0.165509
sd	0.0642413	0.06731	0.0906283	0.0928835
median	0.0320269	0.0402287	0.184976	0.183044
Q25	0.0150715	0.0194139	0.0435081	0.0700715
Q75	0.0901934	0.0926662	0.219943	0.250827
IQR	0.075122	0.0732523	0.176435	0.180756
skew	12.8635	22.4233	1.59106	1.70503
kurt	274.403	634.614	5.3883	5.76912
sp.ent	0.893369	0.892193	0.950436	0.938829
sfm	0.491918	0.513724	0.67547	0.601529
mode	0	0	0.212202	0.267702
centroid	0.059781	0.0660087	0.143659	0.165509
meanfun	0.0842791	0.107937	0.172375	0.185607
minfun	0.0157017	0.0158259	0.0344828	0.0622568
maxfun	0.275862	0.25	0.25	0.271186
meandom	0.0078125	0.00901442	0.79136	0.227022
mindom	0.0078125	0.0078125	0.0078125	0.0078125
maxdom	0.0078125	0.0546875	3.59375	0.554688
dfrange	0	0.046875	3.58594	0.546875
modindx	0	0.0526316	0.311002	0.35
label	male	male	female	female

Table 2: samples

	count	mean	std	min	25%	50%	75%	max
meanfreq	3168.0	0.180907	0.029918	0.039363	0.163662	0.184838	0.199146	0.251124
sd	3168.0	0.057126	0.016652	0.018363	0.041954	0.059155	0.067020	0.115273
median	3168.0	0.185621	0.036360	0.010975	0.169593	0.190032	0.210618	0.261224
Q25	3168.0	0.140456	0.048680	0.000229	0.111087	0.140286	0.175939	0.247347
Q75	3168.0	0.224765	0.023639	0.042946	0.208747	0.225684	0.243660	0.273469
IQR	3168.0	0.084309	0.042783	0.014558	0.042560	0.094280	0.114175	0.252225
skew	3168.0	3.140168	4.240529	0.141735	1.649569	2.197101	2.931694	34.725453
kurt	3168.0	36.568461	134.928661	2.068455	5.669547	8.318463	13.648905	1309.612887
sp.ent	3168.0	0.895127	0.044980	0.738651	0.861811	0.901767	0.928713	0.981997
sfm	3168.0	0.408216	0.177521	0.036876	0.258041	0.396335	0.533676	0.842936
mode	3168.0	0.165282	0.077203	0.000000	0.118016	0.186599	0.221104	0.280000
centroid	3168.0	0.180907	0.029918	0.039363	0.163662	0.184838	0.199146	0.251124
meanfun	3168.0	0.142807	0.032304	0.055565	0.116998	0.140519	0.169581	0.237636
minfun	3168.0	0.036802	0.019220	0.009775	0.018223	0.046110	0.047904	0.204082
maxfun	3168.0	0.258842	0.030077	0.103093	0.253968	0.271186	0.277457	0.279114
meandom	3168.0	0.829211	0.525205	0.007812	0.419828	0.765795	1.177166	2.957682
mindom	3168.0	0.052647	0.063299	0.004883	0.007812	0.023438	0.070312	0.458984
maxdom	3168.0	5.047277	3.521157	0.007812	2.070312	4.992188	7.007812	21.867188
dfrange	3168.0	4.994630	3.520039	0.000000	2.044922	4.945312	6.992188	21.843750
modindx	3168.0	0.173752	0.119454	0.000000	0.099766	0.139357	0.209183	0.932374

Table 3: describe

2.2 Exploratory Visualization

The figure 1 shows the histogram of each feature. The features **IQR**, **meanfun** and **sd** have two peak, which may mean the difference between the voice of female and male.

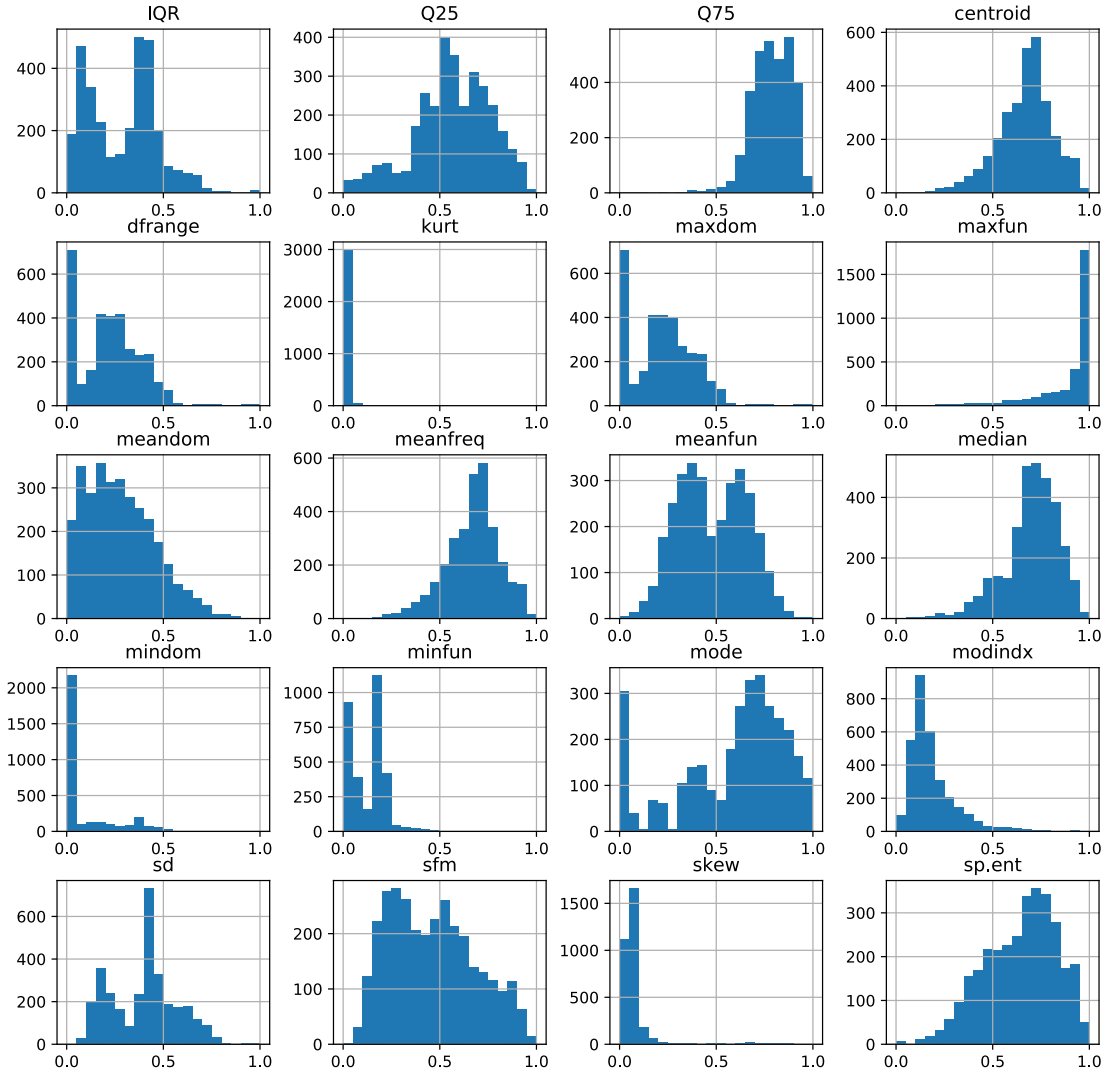

Figure 1: Distribution of Features

References

[1] iFLYTEK. iflytek open platform. http://www.xfyun.cn/.

[2] Kaggle. Gender recognition of voice. https://www.kaggle.com/primaryobjects/voicegender.

[3] Wikipedia. Speech recognition. https://en.wikipedia.org/wiki/Speech_recognition#History.