

男性の声と女性の声の判別

氏名 安藤優希

法政大学情報科学部デジタルメディア学科

学生証番号 15K1103

平成 29 年 1 月 14 日

1 目的・取り組んだ問題

一般的に、私達は男性の声と女性声を聞けば、なんとなく感覚的に性別を判断することができるが、男性と女性の声の違いの境界線はあいまいなものとなっている。私は男性と女性声を数学的にどのように判別したらよいかを本課題で取り組んだ。

$$y = f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

図 1: 多変量正規分布の密度関数

2 方法

判別の方法には、最尤法を用いた。具体的には、男女それぞれ 10 人の 2, 3 秒程度の「あー」という声を用意し、学習データに男女それぞれ 5 人、テストデータに残りの 5 人の声データを使用する。特徴量は周波数の平均値、中央値、短時間エネルギーの三階変化量の最大値、四階変化量の最大値とする。学習データの特徴量の正規パラメーターを推定し、テストデータとなる人の特徴量の多変量正規分布の密度関数 (図 1) を計算する。

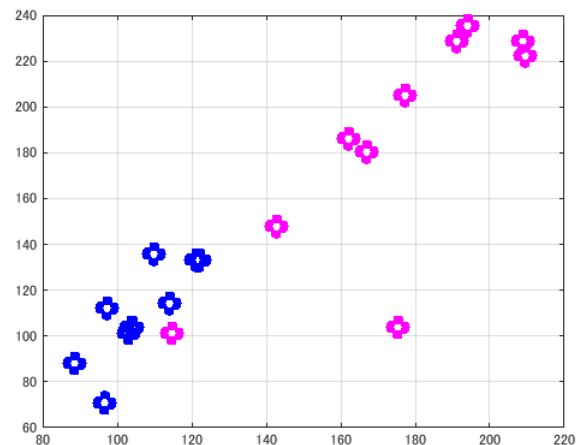


図 2: 声の高さの分類

3 技術詳細

男性と女性の声の違いの判断基準として、音の高さと大きさに注目した。平均値、中央値、最大値、最低値、フレーム間の差分などからうまく分類できる組をグラフにプロットしながら調査した結果、音の高さでは平均値と中央値、音の大きさでは三階変化量の最大値、四階変化量の最大値を特徴量にすることにした。学習データとなる 10 人の声から得た特徴量を並べた行列を男女で二つ作成し、normfit を使用してそれぞれ平均値と共分散を推定する。テストデータとなる別の 10 人の特徴量と先ほど推定した男女の平均、共分散をそれぞれ引数に mvnpdf を使用して多変量正規分布の密度を返す。その密度が大きい方を適切な性別と判別する。

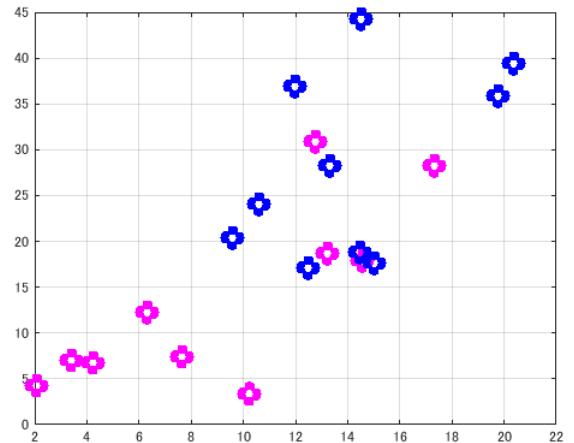


図 3: 短時間エネルギーの分類

4 結果

数通りの学習データとテストデータのグループのパターンを作成し、プログラムを実行した。学習データと

テストデータが似たような特徴量を持つようにグループを分けた場合0.7857と約8割の値を得ることができた。一方、学習データに周波数の高いデータ、テストデータに周波数の低いデータを集めた場合は0.5714と精度が低下した。

5 考察

今回のプログラムでは、学習データとテストデータを均等に分けた場合に約8割という値が得られたため、比較的に良い結果を得ることができた。録音に関してだが、録る環境が様々異なっていたため、いくつかの音声データには雑音が入っていた。改善点として、雑音がなるべく入らないようにとることが挙げられる。また、短時間エネルギーの特徴量に変化量の最大値が有用的だったため、抑揚のない「あー」という声ではなく、アクセントのある「りんご」や「みかん」などの単語を録音したら、より良い精度が得られた可能性があると考ええる。