

メル周波数ケプストラム係数を用いた 音声識別アルゴリズムの提案

溝上 響[†] 亀田 弘之[†] 野村 博^{††}
[†] 東京工科大学コンピュータサイエンス学部
^{††} 国立病院機構 東京医療センター特別研究員

1. はじめに

近年、音声を認識する手法が注目され、身近なところでは、ホームアシスタントや携帯の検索機能などに利用されている。複数の人が同時に会話している中から、発話者ごとの特徴を分離する技術はすでに提案されている。しかし、発話者を推定する人物特定に関しては犯罪捜査以外にはあまり用いられていない。[1] 本稿では、簡便な人物特定のシステムについて提案する。具体的には、ヒトの聴覚特性を考慮した、メル周波数ケプストラム係数(MFCC)に対して Support Vector Machine(SVM)を用いて音声識別を行った。

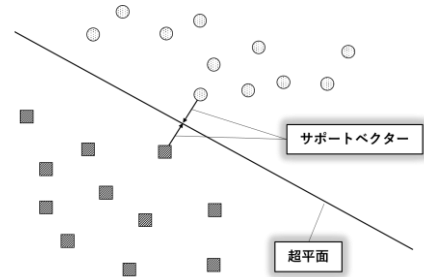


図 2 SVM を説明する図

2. 実験手法

図 1 は、本稿で提案するシステムのブロック図である。Librosa は、音楽・オーディオ分析のために用いられる Python のサポートパッケージで、その中のサブモジュールの一つである feature は、MFCC の算出に用いることができる[2]。

機械学習モデルの一種である SVM は、図 2 のように、データを 2 つのクラスに分類する超平面¹を、サポートベクターを用いて決定する手法である。

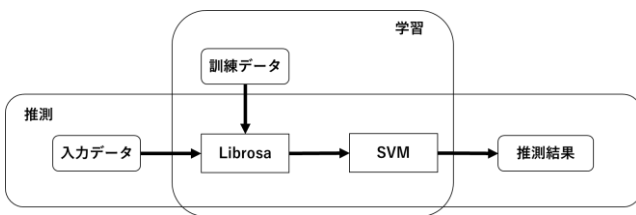


図 1 本システムのブロック図

3. 実験

表 1 に実験条件を示す。被験者 2 人にそれぞれ 4 種類の文章を発話してもらい、その音声データを SVM の学習に用いた。推測には学習時とは異なる文章を用いた。

表 1 実験条件

	発話文章
学習に用いたデータ	あー
	おはようございます
	こんにちは
推測に用いたデータ	ありがとうございます
	お疲れさまでした

4. 実験結果

表 2 に実験結果を示す。今回は、極めて少ない学習データで、正確に音声識別できたことを確認した。

表 2 実験結果

推測に用いたデータファイル	正解	推測結果
hibiki5.wav	被験者A	被験者A
aoi5.wav	被験者B	被験者B

5. おわりに

本稿では、音声データの MFCC を特徴量として、SVM を用い、発話者一人の音声に対する音声識別を行った結果、手法が有効であることを確認した。今後は、被験者を増やした状態で、訓練データの数やパラメータを調整することで精度の向上を目指す。

参考文献

[1] Voice Analysis in Forensics - AzoLifeSciences,
<https://www.azolifesciences.com/article/Voice-Analysis-in-Forensics.aspx>
 [2] 大石 後藤 伊藤 武田, 情報処理学会論文誌, vol. 47, No. 6, pp. 1822-1830,
 June 2006