

# 音声合成可能性を正則化項とする 最小分類誤り学習法の高速化及び実験的評価

岡内 亮太<sup>†</sup> 片桐 滋<sup>†</sup> 大崎 美穂<sup>†</sup>

<sup>†</sup> 同志社大学

## 1. はじめに

音声パターン認識において、有限個の標本のみを用いて分類器を学習する際に起こる問題として、過学習がある。その解決手法として、最近、音声合成可能性を正則化項として用いた最小分類誤り(MCE: Minimum Classification Error)学習法[1]が提案された。本稿では、先行研究で示唆されているベイズ誤り推定能力、及び学習速度の改善について検討する。

## 2. 音声合成可能性を正則化項に用いた MCE 学習法

本手法における音声認識器を図1に示す。入力音声と線スペクトル対(LSP: Line Spectral Pairs)特徴量を用いて、CELPに基づいた音声を合成できる機構を認識器に組み込んでいる。また、クラスモデルは単語単位で構成され、各音素モデルには、マルチプロトタイプの状態遷移モデルを用いている。

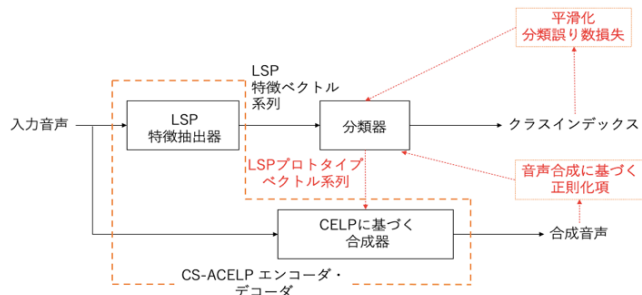


図1. 音声認識器の構成.

## 3. 既存手法の改善

### 3.1 動的時間伸縮法

本手法では、各クラスの識別関数として動的時間伸縮法(DTW: Dynamic Time Warping)を使用している。先行研究のDTWでは、フレーム数×状態数分、経路探索を行っていたが、始点と終点の位置は必ず固定されていることなどから、本手法では、到達不能な経路の削除を行い、DTWを実装した。

### 3.2 プログラミング的アプローチ

本手法の開発言語には、速度と互換性を考慮して、Cythonを用いた。これは、PythonからC言語の利用を可能にしたり、Pythonライクなコードが書ける一方で、繰り返し文がC言語レベルで速く回るといった特徴や、変数の静的型付けが可能であるというような特徴がある。

またコードの可読性は保ちつつ、不要なインスタンス化の削除や、メソッド呼び出しの削減、getter/setterの

みのクラスは構造体を代わりに使用するなどを行った。

## 4. 評価実験

データベースには、男女20人、各話者492単語、総標本数が9840個であるETL-WD-Iを用い、学習用及び試験用にはETL-WD-IをHO法で分割した各4920個の標本を用いた。入力特徴は10次元のLSPとパワー、及びそれらの隣接フレームに対する回帰直線の傾きから成る22次元のベクトルとした。

なお分類誤り率の参考値、ベイズ誤り参考値は、セグメントK平均法を用いて10分割の交差検証(CV: Cross Validation)法で求めた20.04%とする。

### 4.1 速度比較

学習回数は100とし、正則化係数 $\beta$ に6つの値を用いてそれぞれ並列に回し、学習が終わった時間を比較した(表1)。表から、速度に関する大きな進展があったことがわかる。

表1. 新規及び既存コードにおける学習速度比較.

新規コード	既存コード
21:21:13	3days, 7:38:44

### 4.2 各プロトタイプにおける認識率の比較

学習回数100に加えて、プロトタイプ数を3, 5, 7の3通り、正則化係数 $\beta$ に3つの値を用いて学習を行った(図2)。図には、最も優れた、正則化係数が1の場合の結果を示す。なお、赤の破線はCV法による参考値を示している。プロトタイプ数が多い方がCV法によるベイズ誤り値に近い結果が得られた。

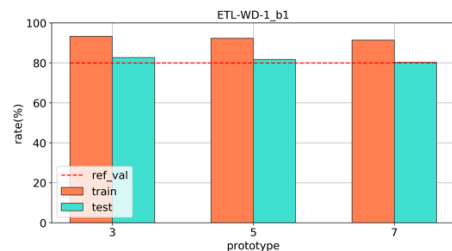


図2. 各プロトタイプ数における正則化係数 $\beta=1$ の認識率.

## 5. まとめ

表1より、手法の改善による速度の向上が見て取れ、学習の高速化による実験効率の向上が期待できる。また図2より、ベイズ誤り状態に近い結果を得るためにはプロトタイプ数の最適化等も重要であることがわかる。

謝辞: 本研究は科研費(18H03266)の支援を受けた。

### 参考文献

[1] 梅崎直統. 同志社大学大学院理工学研究科修士論文, 2019.