ISS-P-061

正則化学習を伴う音声認識器の音声合成能力の実験的評価

大久保　拓海†　　片桐　滋†　　大崎　美穂†
† 同志社大学大学院　理工学研究科

1. はじめに

パターン認識の究極的な目標はベイズ誤り状態の推定である. しかし実際に学習に用いることができる標本数は有限であるので, ベイズ誤りの過小推定, 即ち過学習 が起きてしまう. この過学習問題を回避するために, 先行研究[1] で音声合成可能性を正則化として組み込んだ最小分類誤り(MCE)学習法[2]を用いた分類器が提案され, 過学習抑制効果があることが確認されている. しかし, 音声合成能力そのものや合成音声の質については深く評価されておらず, 分類精度との関連は分かっていない. そこで, 本稿では音声認識器の音声合成能力と認識精度の関連を調べることを目的とする.

2. CELP に基づくプロトタイプを用いた音声合成

音声認識器の音声合成能力を評価するために, 音声符号化方式の一つである CELP[3]を用いた手法を提案する. 通常 CELP では入力音声から計算された線スペクトル対(LSP)と, コードブックから選択された励振信号ベクトルを用いて音声合成する. このとき合成音声が入力音声と近くなるように励振信号は選択される. 提案手法では計算された LSP の代わりに, 認識器のプロトタイプから LSP を抽出し, それを用いて音声合成する. また, プロトイタイプの LSP が音声合成可能な範囲から逸脱しないように, プロトタイプと入力音声の距離を正則化項とし, この正則化項の値が小さくなるように学習を行う.

3. プロトタイプを用いた合成音声の評価法

認識器のプロトタイプを用いた合成音声の音質の指標には, 通常の CELP で合成した音声との波形間誤差を採用する. 誤差の計算は無音区間を除いて行う. 無音区間を除いた認識器からの合成音声$\mathbf{X}=\{x_1,\cdots,x_T\}$, CELP の通常の合成音声$\mathbf{Y}=\{y_1,\cdots,y_T\}$ の波形間誤差 E を次式のように定める.

$$E=\frac{1}{T}\sum_{t=1}^{T}|x_t-y_t|$$

ここで E は 1 単語の誤差である. 評価実験では, 全単語(後述する図 1 及び図 2 中の緑曲線), 正分類した単語(青), 誤分類した単語(赤)の 3 種類の平均誤差を用いる.

4. 評価実験

実験には単語音声データベース ETL-WD-I を用いた. ベイズ誤りの推定値は予備実験より 20.04%とし, これを基準として用いる. 正則化項の係数 β を変化させながら音声

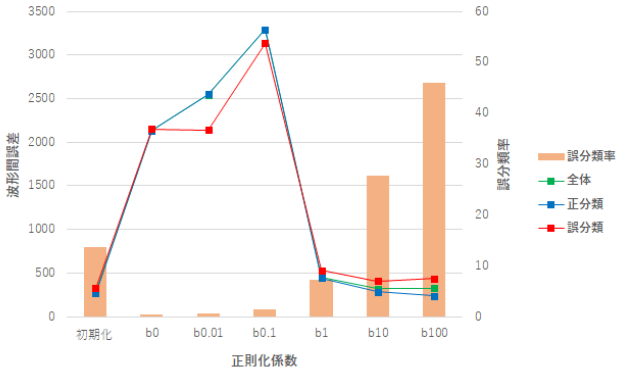

図1. 学習用標本に対する誤分類率と波形間誤差.

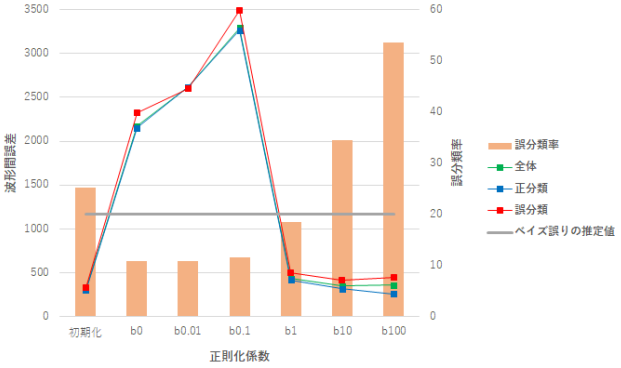

図 2. 試験用標本に対する誤分類率と波形間誤差.

認識・合成実験を行った. 図 1 と図 2 に, それぞれ学習用標本と試験用標本に対する誤分類率と波形間誤差を示す. β が 1 のときに最もベイズ誤りの推定値に近づき, このとき波形間誤差も大きく減少している. このことから音声合成能力の向上とともに, 過学習が抑制されているといえる. β の値が 1 を超えると誤分類率は高くなり, 認識器としての性能は低下してしまうことが分かる.

5. 今後の展望

今後は受聴実験等を通して合成音質の音質を波形間誤差以外の観点から評価していく予定である.

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. 同志社大学修士論文, 2020.
[2] B.-H. Juang, et al.; IEEE Trans. SP, vol. 40, no. 12, pp. 3043-3054, Dec. 1992.
[3] G. 729: 2007.