論文 / 著書情報
Article / Book Information

論題(和文)	大語彙連続音声認識のための言語的音響的属性に基づく単語単位の最適化
Title(English)	
著者(和文)	篠崎隆宏, 古井貞熙
Authors(English)	Takahiro Shinozaki, SADAOKI FURUI
出典(和文)	日本音響学会 2003年春季講演論文集, Vol. , No. 3-4-4, pp. 135-136
Citation(English)	, Vol. , No. 3-4-4, pp. 135-136
発行日 / Pub. date	2003, 3

大語彙連続音声認識のための言語的音響的属性に基づく単語単位の最適化 *

◎篠崎 隆宏　古井 貞熙　(東工大)

1　はじめに

日本語は単語の単位が明確ではないことから、言語モデルの作成において先ず単語単位の設計が必要となる。Ngram モデルにおいては長い単語単位を用いることにより、広い言語コンテキストをモデル化することが出来るが、学習データが有限の場合モデルの正確な推定が難しくなる。また実験的には短い単語は認識誤りの原因になりやすいことが知られている [1]。音声認識処理の際には言語重みや単語挿入ペナルティを設定することが一般的であるが、このようなパラメータや言語モデルの推定精度を考慮しながらシステムの認識性能を単語単位の関数として理論的に解析することは困難である。

そこで本稿では、まず個々の単語の単語長や出現回数が認識難易度に及ぼす影響を実験的に分析しモデル化を行い、得られた認識難易度モデルに基づいて評価値を最大とするように単語を順次併合する、単語単位の最適化手法を提案する。これまでに相互情報量や単語対頻度のみに基づいた最適化手法は提案されていたが、最適化基準が認識率ではないことが欠点であった。本手法は単語長と出現回数を同時に考慮することが可能であり、さらに認識率を最適化基準とする手法である。またコンテキスト長とモデル推定精度のバランスの観点からは可変長 Ngram が研究されているが、デコーディング処理が複雑になる欠点がある。本手法では単語単位が決定された後は、通常のデコーダがそのまま使える利点がある。日本語話し言葉コーパス CSJ の学会講演を用いた認識実験において、本手法が有効に働くことを確認した。

2　単語正解確率モデル

単語の長さ、出現回数と認識難易度の関係を分析しモデル化するために、各単語毎に以下の変数を設定した。

Cor: 単語正解率 (%) (0 or 100)

NP: 単語の音素数

WF: 単語の学習セット中での出現回数

LF: 出現回数 *WF* の対数：$\log_{10}(WF + 10^{-6})$

Cor は各単語に対して 100(正しく認識された) か 0(間違って認識された) かどちらかの値をとる。間違いとするのは他の単語と置き換わって認識された場合（置換誤り）および対応する単語が出力されなかった場合（削除誤り）で、認識出力のみが存在し対応する正解単語が存在しない場合（挿入誤り）は本研究では対象としなかった。単語正解確率モデルを単語正解率 Cor の、単語長 NP と対数出現回数 LF の条件付確率として式 (1) のように定義する。

$$P\left(Cor = 100 | NP, LF\right). \qquad (1)$$

3　単語単位の最適化

単語対の併合に伴い単語の定義が変化するため、システムの認識率評価には文字正解率を用いることにする。またここでは文字正解率の近似値として、その下限値を推定し用いることにする。システムの文字正解率の下限値は α を単語の文字数として、単語正解確率モデル (1) を用いて式 (2) により見積もることが出来る。

$$E\left[Cor\right] = \frac{\sum_w P\left(Cor = 100 | NP, LF\right) \cdot WF \cdot \alpha}{\sum_w WF \cdot \alpha}. \qquad (2)$$

単語が誤認識される場合、その単語を構成する全ての文字が間違いであるならば、このようにして求められた認識率は文字認識率となる。またこのようにすることで、以下に示す評価関数を高速に計算することが可能となる。

単語対 $\langle w_1, w_2 \rangle$ の併合により新しく生成された単語 $w_{1,2}$ の各属性値は次のように求める。

- $NP\left(w_{1,2}\right) = NP\left(w_1\right) + NP\left(w_2\right)$
- $WF\left(w_{1,2}\right) = w_1$ と w_2 が学習セット中でこの順に並んでいる回数

また w_1, w_2 の出現回数も、併合の影響を受ける。併合後の語彙中の w_i $(i = 1 \text{ or } 2)$ を $w_i^{'}$ とすると、新しい属性値は以下のように求めることが出来る。

- $WF\left(w_i^{'}\right) = WF\left(w_i\right) - WF\left(w_{1,2}\right)$

単語の併合操作を行う前と後で言語モデルを作成し使用した場合の、認識システムの認識率の差の推定値を評価関数とする。単語の併合を評価値の合計が最大となるように繰り返すことで音声認識に適した認識単位が得られると期待できる。単語の併合を繰り返す場合、併合の順番は全体としての評価値に影響を及ぼすが、計算量の点から貪欲法を採用した。最適化手順を以下に optwordunit() としてまとめる。ループ中で生成された併合単語も以後の併合候補に含まれる。学習セットに実際に登場する単語対の種類は限られていることなどを利用することで高速に併合単語を求めることが可能である。

```
procedure optwordunit() {
 for i = 1:maxiter {
   select word pair <w1,w2>
     which maximizes delta(w1,w2);
   merge_all_adjacent(w1,w2);
 }
}
```

∗ A lexicon optimization method for LVCSR based on linguistic and acoustic characteristics of words

By Takahiro Shinozaki and Sadaoki Furui (Tokyo Institute of Technology)

4　実験結果

言語モデルの学習セットはCSJの学会講演と摸擬講演を合わせた610講演である。形態素解析には、NTTで開発された形態素解析ツールJTAGを使用した。この単位を使用して学習したモデルをベースラインモデルとする。言語モデルに使用した語彙は学習セット中で出現頻度の高い30k単語である。2gramと逆向き3gramを作成し、認識処理にはJuliusを用いた。音響モデルは男性話者による学会講演59時間を用いて学習した。

単語正解確率モデルの学習のため、開発セットとしてCSJ中の男性話者による44学会講演を対象にベースラインモデルを用いて認識を行った。ロジットモデルとしてモデル化した単語正解確率モデルを式(3)に示す。

$$\begin{aligned} &P\left(Cor = 100|NP, LF\right) \\ &= \Lambda\left(0.70NP - 0.03NP^2 + 1.60LF\right. \\ &\quad \left. -0.12LF^2 - 5.17\right). \end{aligned} \tag{3}$$

ここでΛはロジスティック関数であり、次式で表される。

$$\Lambda\left(x\right) = \frac{e^x}{1 + e^x}. \tag{4}$$

図1に単語長NPで層別した対数出現回数LFと単語正解率Corの関係を示す。図で実線は式(3)により計算した値、破線は実測値である。

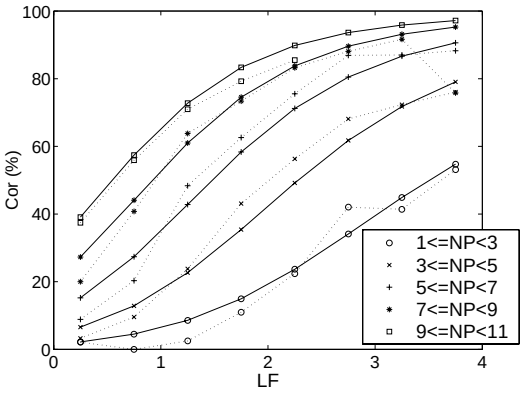

図 1. Log word occurrence count and recognition correctness.

アルゴリズムoptwordunitを用い、学習セットに対し、単語単位の最適化を行った。読点は併合の対象としなかった。初期状態での異なり単語数は約35kである。併合は500種類の併合単語が得られるまで繰り返した。この操作により学習セットの単語数は最適化前の1.5Mから1.3Mに減少した。単語対の併合による評価値の累積値は1.39%であった。評価値、およびその累積値の推移を図2に示す。

最適化により得られた学習セットを用て単語単位最適化言語モデルをベースラインモデルと同様に作成した。単語対の併合により学習セットに含まれる語彙数が変化するが、言語モデルに使用した語彙は最適化後の学習セットにおける出現頻度の高い30k単語である。

認識結果の比較は文字正解率および文字正解精度を用いて行った。提案手法の評価に用いたテストセットはCSJ中の男性話者による10学会講演である。認識実験を行う際の言語重みと挿入ペナルティは、ベースラインシステムと単語単位最適化システムそれぞれにおいてテストセット全体の文字正解精度が最大となるように選んだ。結果を図3に示す。

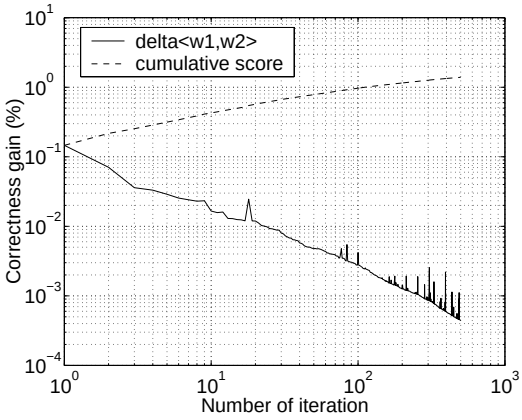

図 2. Changes of the evaluation and accumulated values in 500 iterations.

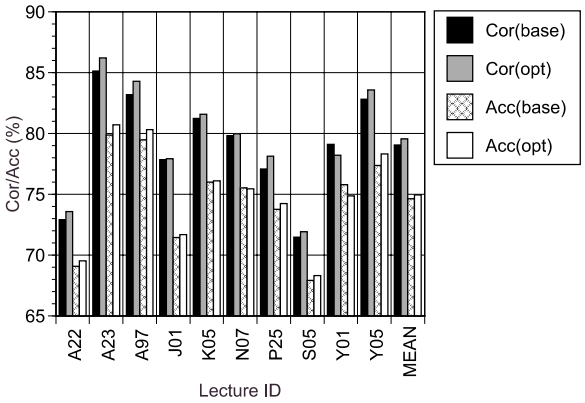

図 3. Character correctness and accuracy before (*base*) and after (*opt*) the optimization.

文字正解率は10講演中9講演で、文字正解精度は8講演で改善が見られた。10講演の平均では正解率と正解精度はそれぞれ0.48%、0.33%向上した。

見積もられた文字正解率の改善1.39%と比べると、小さな改善となったが、この原因としては単語単位最適化アルゴリズムで使用している単語正解確率の予測誤差などが考えられる。

5　まとめ

データから学習した単語正解確率モデルを用いることでシステムの認識率を評価関数として使用する、単語単位の最適化手法の提案を行った。日本語話し言葉コーパスを用いた認識実験により本手法が認識性能の向上に有効であることを示した。

参 考 文 献

[1] 篠崎 隆宏, 古井 貞熙, “話し言葉認識における決定木を用いた誤り要因の分析,” 音講論, 1-1-9, pp.17-18, (2001-10).