論文 / 著書情報
Article / Book Information

論題(和文)	形態素の読みの確率を考慮したニュース音声のディクテーション
Title(English)	
著者(和文)	高木幸一, 古井 貞熙
Authors(English)	Koichi Takagi, SADAOKI FURUI
出典(和文)	日本音響学会 1998年春季講演論文集, Vol. , No. 1-6-5, pp. 9-10
Citation(English)	, Vol. , No. 1-6-5, pp. 9-10
発行日 / Pub. date	1998, 3

1—6—5

形態素の読みの確率を考慮したニュース音声のディクテーション*

◎高木 幸一　古井 貞熙 (東工大)

1 はじめに

日本語独特の現象として,1つの形態素に対する読みの多様性が挙げられる. 例えば「円」には"e-N","m-a-r-u"などの読みが挙げられる. ところが,"m-a-r-u"は実際にはほとんど使われない読みにもかかわらず,テキストレベルでこの「円」と言う単語が頻出するために,言語モデル上,高いスコアが与えられてしまう. そこで本稿では,これにより生ずる認識性能の劣化を抑えるため,発音辞書の各形態素の読みごとに確率を付与する方法を提案する. またニュース音声のディクテーションの高度化のためのその他の種々の試みについても報告する.

2 言語スコア計算の近似

2.1 読みの確率の近似

単語 w_k に対し,読みの候補が $r_k = r_{k1}, r_{k2}, \ldots$ とあるとする. 日本語の場合,読みが違ってしまうと別の意味の単語となってしまうパターンは数多い. 当然,読みにより前後に出てくる単語も違ってくる. そこで本来なら単語列生成確率として読みごとに

$$P(w_{k=1}^{n}(r_k)) = \prod_{k=1}^{n} P(w_k(r_k)|w_{i=1}^{k-1}(r_i)) \quad (1)$$

を計算すべきであるが,単語が細分化されるため,データがさらにスパースになってしまう. そこで class-Ngram の考え方を応用し,

$$P(w_k(r_k)|w_{i=1}^{k-1}(r_i)) \approx P(w_k(r_k)|w_k)P(w_k|w_{i=1}^{k-1}) \quad (2)$$

で近似する. 現在まで行われてきた方法は式(2)の右辺第1項が1と固定されていたものに相当する.

具体的には発音辞書の各読みがどの程度の割合で出て来るかを言語DBから算出し,その確率を発音辞書の各読みに付与した. そして,その確率 P_{prob} にある重み r をつけ (P_{prob}^{r}),これを式(2)の第1項として使用した. 実際には確率値は対数領域で計算されるため,$r \log P_{prob}$ を各単語モデルの遷移の際に加えることになる.

2.2 新語登録

連続音声認識において,未知語を正しく検出することは現時点では極めて難しい. そこで,ニュース音声のディクテーションにおいては,直前までのニュースや新聞のテキストを用いて未知語を検出し,それをマニュアルで新語 u として読みとともに登録することが現実的な解であると考えられる. この際,その新語 u の言語モデルを次のように与える方法を検討した.

$$P(u_k|w_{i=1}^{k-1}) \approx P(u_k|C_k)P(C_k|w_{i=1}^{k-1}) \quad (3)$$

$$\approx \frac{1}{K}P(C_k|w_{i=1}^{k-1}) \quad (4)$$

ここで,C は全新語のクラス,K は学習テキスト中の未知語の種類数である.

3 言語モデル

言語モデルを,時期の違いによる性能を見るために,NHK ニュース原稿データベースの 1992 年 8 月から 1996 年 5 月までの約 5 年間の原稿 [1](以下「テキスト DB」と言う) の全部または一部から別々に作成した.

語彙の単位は「形態素」とし,解析ツール Chasen [5] を用いて解析した. なお,誤読を少しでも解消するために,漢数字に関してのみ単位ごとに区切るよう (例えば「八千三百四十七」→「八千」「三百」「四十」「七」) 辞書を改良した. さらに,語彙サイズはカバー率が約 98%となるよう,出現頻度に従い上位 20k を選択した.

また,言語モデルの作成には CMU/Cambridge Toolkit ver.2.03 [6] を用いた.

4 認識実験

4.1 分析, 音響モデル

音響特徴量は 16 次の LPC ケプストラムと正規化対数パワー,及びそれらの 1 次時間微分の計 34 次元を用いた. 音響モデルは表 1 の男声データを用い, Tree-based clustering による状態共有化を行った HMM [4] を採用した. 構築されたモデルの総状態数は 2106, ガウス分布の混合数はすべて 4 である.

表 1: 音響モデル学習用 DB

ASJ 音素バランス文連続音声 DB
ASJ 案内タスク連続音声 DB
ATR B セット (計 53 話者, 13270 発話)

4.2 評価用音声

音声認識に用いるニュース音声の評価セットは,96 年 6 月に実際放送されたものの中の,スタジオ内で話された比較的 clean な音声 (男声) を用いた. さらにそれを話者がメインのアナウンサー (anchor set) か否か (others set) に分類した. 表 2 にその概要を示す.

*Dictation of broadcast news speech using word pronunciation probability.
By Koh'ichi Takagi and Sadaoki Furui (Tokyo Institute of Technology)

anchor set は時事的なニュースを多く含むせいか、時期が離れるごとに未知語率が上がっていることがわかる．また,others set の未知語率は極めて大きいことがわかる．

表 2: 評価用音声データの未知語率とテキスト DB

		anchor	others
話者数		5	6
総発話数		100	125
総単語数		3957	2146
♣ 92 年 7 月-96 年 5 月	(22.0M)	0.81%	3.40%
95 年 5 月-96 年 5 月	(9.4M)	0.78%	3.49%
94 年 4 月-95 年 9 月	(9.2M)	1.06%	3.08%
92 年 7 月-94 年 9 月	(9.7M)	1.36%	3.31%

下 4 行はテキスト DB の期間 (量) と
各 set に対する未知語率を表している

4.3 認識実験結果

まず,テキスト DB を時期により分けて学習した言語モデル (bigram) を用いてその Test set perplexity を求めた．また,認識システムにその言語モデルを適用したときの性能を調べた．結果を図 1 に示す．時事的なニュースの多い anchor set は未知語率同様，認識性能,test-set perplexity ともに評価用音声にテキスト DB の時期が近づく程良くなる傾向にあることがわかる．

次に，第 2.1 節に挙げた方法で重み r の値を適当に変化させ，読みの確率を考慮した発音辞書を評価用音声に適用した．言語モデルはテキスト DB(表 2 の♣のデータ) から作成した bigram を使用した．その結果,$r = 5$(anchor set) の条件で誤り率が 5%ほど減少した (表 3)．

図 1: テキスト DB の時期の違いによる単語正解精度

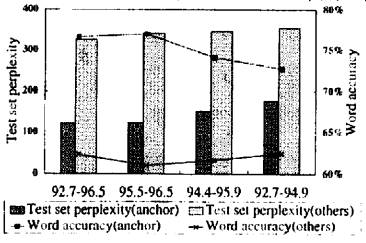

表 3: 読みを考慮した認識実験結果 (単語正解精度)

r	anchor set	others set
0(normal)	**76.6%**	**62.3%**
0.5	76.9%	62.2%
1	77.3%	61.9%
2	77.5%	62.5%
5	**77.9%**	63.4%
7	77.5%	**63.5%**
10	77.0%	62.7%

表 4: 新登録語認識実験結果 (単語正解精度)

	anchor set		others set	
	20k 語	+登録語	20k 語	+登録語
登録語 → 登録語	0	22	0	56
登録語 → 他	32	10	73	17
他 → 登録語	0	9	0	20
単語認識精度	76.6%	77.6%	62.3%	66.3%

また，第 2.2 節に挙げた方法で言語モデルに新登録語クラスを追加し，新登録語に対する認識実験を行ったところ,7 割近い (anchor set=22/32=69%, others set=56/73=77%) 新登録語を正しい新登録語と認識することができた (表 4)．

5 おわりに

日本語大語彙連続音声認識を難しくしている原因の一つと考えられる形態素の読みの多様性に対する配慮として，発音辞書の各読みに確率を付与する方法を提案した．これをニュース音声のディクテーションシステムに適用したところ，誤り率を最大 5%ほど削減することができた．

その他，新語登録による認識性能の向上，時期が違う DB から作られた言語モデルに対する認識性能の違いも確認することができた．

今回読みの確率を算出する際に使用したデータは比較的小さいものであるため，より多量のデータを用い，読みの精度を上げることにより，一層の性能向上が期待できる．また，読みの確率のタスク適応も必要になってくるものと思われる．

謝辞

NHK ニュース原稿及び音声データを提供して頂いた NHK 放送技術研究所，音響モデルの構築に尽力して頂いた山形大の堀貴明氏，日頃御指導いただく NTT ヒューマンインターフェース研究所の大附克年氏，松永昭一氏に感謝します．

参考文献

[1] 安藤，他，“ニュース音声データベースの構築”，春季音学講論，1997，pp. 157-158

[2] 田口，他，“ニュース音声を対象とした大語彙連続音声認識”，春季音学講論，1997，pp. 65-66

[3] 小林，他，“ニュース音声認識システムの検討”，秋季音学講論，1997，pp. 103-104

[4] 大附，他，“ニュース音声を対象とした大語彙連続音声認識と話題抽出”，信学技報，1997，SP97-27

[5] http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html

[6] http://svr-www.eng.cam.ac.uk/~prc14/toolkit.html

[7] Steve Young, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland, “The HTK Book(for HTK version 2.1)”,Cambridge Research Lab., Mar. 1997