

音声対話システムにおけるバイグラムの タスク適応化の検討

堀 賢史 小暮 悟 中川 聖一

豊橋技術科学大学 情報工学系

1 はじめに

ある新しいタスクの対話音声認識するためのバイグラムを構築するには、そのタスクの対話文集合が数千文は必要となるが、その文集合を収集するのは困難である。[1] そこで、あるタスクの対話文集合数百文とそれに関連したタスクの文集合からバイグラムを構築する方法が考えられる。本研究は、このようなバイグラムの構築法を検討し、文献検索のタスクにおいて、パープレキシティと音声認識率の評価を行なった。

文献検索対話文からユーザーの発話を、関連したタスクとして 11 種類の対話文集合 [3][4] を使用し、文献検索対話文のみで学習文数を変えた言語モデル、関連したタスクの文集合のみの言語モデル、両者を混合した言語モデルで、認識率等を求め、混合した言語モデルについては、MAP 推定による適応化を行なった。また、未知の固有名詞を語彙に追加した場合の方法を検討した。

2 言語モデルの構築と認識実験

2.1 言語モデルの構築

言語モデルの構築に共通の条件は次の 3 点である。

- 「JUMAN3.2」で形態素解析を行なう
- バイグラムの cut-off は 0 とする
- 語彙数の制限はしない

文献検索対話文のユーザーの発話 495 文を、学習文 395 文 (3352 語)、テスト文 100 文 (818 語) に分割し、学習文 395 文中、100 文、200 文、395 文で学習した 3 つの言語モデルを構築した。

関連したタスク (複数タスク) は、道案内、スケジュールリングなどの 11 種類の対話文で、合計 8927 文を使用し、言語モデルを構築した。

また、この両者のコーパスを混合した言語モデル (mix) として、複数タスク全文と、文献検索対話文の学習文から 200 文を使用し、この 200 文の相対頻度を高めるために、5 倍、10 倍の重みを与えて言語モデルを構築した。また、比較のためにニュース文約 47 万文から語彙数 20000 の言語モデルも構築した。

A study on bi-gram adaptation to a new task for spodialogue systems

Kenji Hori, Satoru Kogure and Seiichi Nakagawa
Toyohashi University of Technology, Department of
Information and Computer Sciences

それぞれの言語モデルの語彙数、パープレキシティとカバーレージを表 1 に示す。ここで補正パープレキシティ (APP) は未知語の確率を $\frac{1}{50000}$ としている。混合した言語モデルは文献検索対話文の 395 文で学習した言語モデルと同等のカバーレージを持っているが、パープレキシティはその 2 倍近くであることが分かる。

表 1: 各言語モデルによる文献検索対話文のテスト文 (100 文、818 語) の評価

言語モデル	語彙数	PP	APP	カバー率
文献検索 100 文	259	22	83	82.9 %
文献検索 200 文	373	20	45	89.5
文献検索 395 文	569	18	29	94.0
複数タスクのみ	3607	65	231	81.0
mix 5 倍	3766	36	55	93.9
mix 10 倍	3766	32	51	93.9
ニュース文	2 万	880	1766	86.5

2.2 音声認識実験

前節の各言語モデルを用いて、音声認識実験を行なった。男性 3 名から収録した文献検索テスト文 100 文の音声を認識した。なお、音響モデル、特徴パラメータ等は表 2 のものを使用している。単語認識率を表 3 に示す。ACC. は挿入誤りも考慮した単語認識精度、COR. は単語正解率である。

文献検索対話文 395 文での認識率が最も良く、文献検索対話文 200 文で作成したモデルと文献検索と複数タスクを混合し 10 倍で重み付けしたモデルがほぼ同じ認識率であった。

表 2: 音響モデルと特徴パラメータの条件

サンプリング周波数: 12kHz
窓関数: 21.33ms ハミング窓
フレーム周期: 8ms
分析: 14 次元 LPC 分析
音響モデル:
音節モデル, 6 状態 5 出力分布, モデル数 114
離散継続時間付き連続出力分布型 HMM
4 混合ガウス分布, 全共分散行列
特徴パラメータ:
LPC_MEL_CEP (10 次×4 フレームを 20 次元に KL 展開で圧縮) + ΔCEP (10 次) + ΔΔCEP (10 次) + Δ POW + ΔΔPOW

表 3: 7つの言語モデルでの認識実験の結果

言語モデル	COR.(%)	ACC.(%)
文献検索 100 文	57.7	46.0
文献検索 200 文	65.4	57.0
文献検索 395 文	71.7	65.9
複数タスクのみ	49.0	42.1
mix 5 倍	60.5	53.2
mix 10 倍	64.9	56.9
ニュース文	30.0	21.2

3 MAP 推定による適応化

両コーパスを混合した言語モデルについて、関連したタスク 11 種類から作成した言語モデルを標準言語モデルとし、文献検索対話文 50、100、200 文を適応化サンプルとして与え、パープレキシティ、単語認識率の変化を調べた。(1) 式に $\alpha = 0.3$ を与えて MAP 推定 [2] を行なった結果を表 4 に示す。適応化サンプル数が増えるほど、COR. と ACC. は大きくなり、パープレキシティは減少することが分かる。

$$prob = \alpha \cdot prob_0 + (1 - \alpha) \cdot prob_1 \quad (1)$$

$prob$... MAP推定後の条件確率
 $prob_0$... 標準言語モデルでの条件確率
 $prob_1$... 適応化サンプルでの条件確率
 α ... 重み

表 4: MAP 推定 $\alpha = 0.3$ の時の単語認識率

追加文数	COR.(%)	ACC.(%)	PP	APP
0	49.0	42.1	65	230
50	56.7	50.5	48	77
100	61.2	55.4	39	63
200	64.6	58.5	28	46

4 語彙への固有名詞の追加

各言語モデルのテスト文 (818 語) に対する未知語総数、未知語の種類数、未知語中の固有名詞の数を表 5 に示す。

これを見ると、言語モデル mix の未知語総数が文献検索 395 文で作成した言語モデルと同程度であるのに、未知の固有名詞が文献検索 200 文の場合と変わっていないことから、複数タスクでは、目的のタスクの固有名詞をカバーするのは困難であるといえる。

また、未知の固有名詞が未知語総数のおよそ 1/4 を占めており、認識に大きな影響を与えていると考えられる。そこで、文献検索 200 文から作成した言語モデルについて未知の固有名詞を語彙に追加し、そのユニグラムを 1/2000 とした場合と、固有名詞に関してクラスバイグラムを採用しクラスから各固有名詞への確率を 1/100 とした場合について認識実験を行なった。表 6 に結果を示す。

固有名詞を登録することにより ACC. は 8~9% 上昇し、文献検索 395 文の結果に匹敵する単語認識率が得られた。このことから、目的のタスクに依存した固有名詞を認識できるかどうかの結果に大きく影響することがわかる。

表 5: 各言語モデルでのテスト文の未知語数

言語モデル	未知語		
	総数	種類数	固有名詞
文献検索 100 文	147	120	31
文献検索 200 文	101	90	27
文献検索 395 文	58	53	14
複数タスクのみ	160	91	43
mix 5 倍	60	51	27
mix 10 倍	60	51	27
ニュース文	128	78	38

表 6: 固有名詞の追加実験結果

追加モデル	COR.(%)	ACC.(%)
ユニグラム	69.1	65.1
クラスバイグラム	69.6	66.3

5 まとめ

以上の実験より現段階では、文献検索対話文の学習文全文を使用した言語モデルで最も良い単語認識率が得られるという結果となった。

また、未知の固有名詞を語彙に追加することで認識率を向上させることができた。このことから、MAP 推定を行なう言語モデルと各タスクに依存した固有名詞のクラスバイグラムを融合することで、各種の対話文での認識率を向上できると考えられる。

今後は、固有名詞のクラスの細分類化と単語バイグラムとの融合、および、言語モデルの適応化を検討していく予定である。

参考文献

- [1] 中川, 大谷:「Bigram の使用による話し言葉用確率文脈自由文法の自動学習」, 情報処理学会論文誌, Vol.39, No.3(1998.3)
- [2] 中川, 赤松, 西崎:「音声認識用言語モデルのためのタスク適応化と定型表現の利用」, 言語処理学会, 自然言語処理 Vol.6, No.2, pp.97-115(1998)
- [3] 文部省重点領域研究 [音声対話], CD-ROM Vol.1(1994), Vol.2-4(1995)
- [4] 電総研道案内対話音声コーパス, CD-ROM Vol.1-7(1998)