

CFG/bigram を使用した対話音声認識における
意味理解の比較検討

4 R - 4

小暮 悟¹ 伊藤 敏彦¹ 廣瀬 良文² 甲斐 充彦¹ 中川 聖一¹

豊橋技術科学大学 情報工学系¹

奈良先端科学技術大学院大学 情報科学研究科²

1 はじめに

本研究室で開発中の富士山観光案内システム [1] は、音声認識に用いる言語的な制約として文脈自由文法 (CFG) を採用してきた。

今回、これまでの評価実験で得られている対話コーパスから単語の bigram を求め、その bigram を言語的制約に用いる認識器を構築した。次に、上記とは異なる音声対話コーパスを用いて、CFG 及び bigram の認識器による認識率を比較した。また、それぞれの認識結果を既存の意味理解部に入力し、それぞれの意味理解率を比較検討した。

本稿では、これらの結果について報告する。

2 CFG/bigram を用いた音声認識

本研究室の富士山観光案内システムにおいて被験者実験を行なった時の対話音声収録し、これを使用して認識実験を行なった。

今までに 3 回の被験者実験によって、計 2500 発話の対話音声を得られている [2, 3, 1]。これらの発話のうち、2063 発話、12,763 単語を bigram を構築するためのトレーニングデータとして使用し、残りの計 437 発話、2592 単語をテストデータとして使用した。テストデータの平均文長は 5.9 単語/文であった。

CFG(語彙数 359 単語) と条件を合わせるために語彙数 359 単語の単語 bigram を学習した [4]。表 1 に学習した bigram 言語モデルのパープレキシティを示す。bigram 言語モデルのテストデータに対する補正パープレキシティは 9.3 であった。一方、人手で作成した CFG(文法規則数は 501 規則) では 437 文中 274 文を受理できて (受理率は 62.7%)、受理できた文に対するパープレキシティは 136 であった (表 2)。

パープレキシティの値が大きいのは、CFG の文法が確率を用いず、助詞落ちや倒置表現も受理するように作成されているのも一因である。しかしながら、実際の発話には倒置表現はなかった。

表 3 に CFG と bigram の単語認識率と文正解率を示す。これより、bigram 言語モデルを使用の方が CFG の場合と比較して、単語認識率で 18%、文正解率で 29%、助詞誤りを許した文正解率で

12% 良い結果を得ていることが分かる。CFG のパープレキシティが 136 と bigram と比較して非常に大きいことから、bigram 言語モデルを用いた認識器が、CFG を用いるものよりも認識率が良くなっていることが言える。例えば図 1(1) に示すように、CFG では大きな誤りが生じることがある。

表 1: bigram のパープレキシティと受理率

学習データ	文数	2063	
	未知語数 [総数]	50[225]	
	未知語率	1.5%	
	カバー率	bigram	97.1%
		unigram	98.5%
	パープレキシティ	4.0	
	補正パープレキシティ	4.4	
テストデータ	文数	437	
	未知語数 [総数]	10[56]	
	未知語率	1.8%	
	冗長語率	0.2%	
	カバー率	bigram	89.9%
		unigram	98.2%
	パープレキシティ	8.7	
	補正パープレキシティ	9.3	
	文の受理率	88.3%	

表 2: CFG のパープレキシティと受理率

学習データ	文数	2063
	パープレキシティ	129.4
	文の受理率	68.3%
テストデータ	文数	437
	パープレキシティ	135.6
	文の受理率	62.7%

しかし、すべてにおいて bigram の方が良いと言うわけではない。例として、入力文「ペンションマリエは朝食は付きますか」に対する CFG の認識結果と bigram の認識結果を図 1 (2) に示す。この例では、CFG で完全に認識されている文が、bigram を用いた場合には、「付く」が欠落している。bigram の学習に使用した対話コーパス中に「付く」という単語が存在しなかったためである。

このような比較的小規模なタスクでも、学習データが 1000~2000 文程度あれば、CFG よりも良い音声認識用の言語モデルを作成することが出来る [5]。

Comparison of semantic interpreter on spoken dialogue using CFG/bigram based speech recognizer

Satoru Kogure, Toshihiko Itoh, Yoshifumi Hirose, Atsuhiko Kai and Seiichi Nakagawa

Toyohashi University of Technology, Department of Information and Computer Sciences

表 3: CFG と bigram の単語認識率と文正解率

上段：単語認識率 (accuracy)[%]
 中段：完全文正解率 [%]
 下段：助詞誤りを許した文正解率 [%]

話者	HB	ST	TY	YY	平均
CFG	69.4	79.3	78.4	59.2	73.3
	15.2	53.2	53.6	17.3	39.3
	78.8	69.4	92.7	78.8	76.9
bigram	92.2	90.5	93.0	90.7	91.5
	55.2	65.2	90.9	61.3	69.1
	92.9	76.6	100.0	93.3	88.6

(1) 入力文：
 西湖でテニスはできますか
 CFG の認識結果：
 どれ西湖で出来ますか
 bigram の認識結果：
 西湖でテニスはできますか

(2) 入力文：
 ペンションマリエは朝食は付きますか
 CFG の認識結果：
 ペンションマリエは朝食は付きますか
 bigram の認識結果：
 ペンションマリエは朝食はますか

図 1: CFG と bigram による認識結果例

3 CFG/bigram を用いた音声認識結果の意味理解

前節で述べた認識結果を既存の意味理解部に入力し、CFG と bigram の違いによる意味理解率を調べた。意味理解率は、意味理解部が出力する意味表現 (図 2) を人間が見て、意味理解の可否をチェックすることで求めた。「応答文が正しく生成できたか？」という尺度もあるが [6]、今回は調べていない。

表 4 に今回の実験結果による意味理解率を示す。CFG に対する意味理解率は 57.0% で、bigram に対する意味理解率は 68.4% であった。前回の実験

(1) 入力文：
 ペンションマリエは朝食は付きますか
 意味表現：
 (付く (FORM YN-Q) (NEGATION NIL)
 (AT-LOC (ペンションマリエ))
 (OBJ (朝食)))

(2) 入力文：
 富士大科学館の入場料はいくらですか
 意味表現：
 (かかる (FORM WH-Q) (NEGATION NIL)
 (TARGET (COST))
 (OBJ (入場料 (OF (富士大科学館))))
 (COST (WH (AKO (COST))))))

図 2: 意味表現

において、CFG の認識結果文の意味理解率は 4 人の平均で 55.4% (全被験者の平均で 54.0%) である [3]。bigram では認識結果文の意味理解率が 70% 近くまで向上した。なお、書き起こし文を入力した時の意味理解率は 78.3% であった。

表 4: 意味理解率

上段：意味理解できた文数
 下段：意味理解率 [%]

話者	HB	ST	TY	YY	合計
文数	85	167	110	75	437
CFG	39	96	76	38	249
	45.9	57.5	69.1	50.7	57.0
bigram	63	98	84	54	299
	74.1	58.7	76.1	72.0	68.4
書き起こし文	70	113	91	68	342
	82.4	67.7	88.2	90.7	78.3

4 まとめ

今回、bigram を用いた認識器を使用した方が CFG を使用するよりも、音声認識率、意味理解率ともに高いことを示した。

今後は、bigram を組み込んだ認識器を用いて、観光案内のオンラインによる被験者実験を行ない、実際に CFG と bigram による認識、意味理解の違いを比較していく必要がある。

参考文献

- [1] 中川, 傳田, 伊藤:「マルチモーダル観光案内対話システム」, 人工知能学会誌, Vol.13, No.2 (1998.3).
- [2] 傳田, 伊藤, 中川:「マルチモーダルインターフェースを備えた観光案内対話システムの評価」, 人工知能学会全国大会, 15-09, pp.431-434(1996)
- [3] 傳田, 伊藤, 小暮, 中川:「マルチモーダルインターフェースを備えた観光案内対話システムの評価実験」, 情報処理学会, SLP-15-8, pp.47-52,(1997)
- [4] 廣瀬, 甲斐, 中川:「バイグラム言語モデルに基づく対話音声認識における冗長語・未知語処理」, 信学技報, SP98-4, pp.25-32,(1998.4)
- [5] 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)
- [6] 伊藤, 小暮, 中川:「強制的応答を備えた音声対話システムとその評価」, 情報処理学会論文誌, pp.1248-1257,(1998.5)