自然言語処理 137−11
音声言語情報処理 31−6
(2000. 6. 2)

音声対話システムのための未知語の登録を考慮した言語モデルの構築

小暮悟 堀賢史 中川聖一

豊橋技術科学大学 情報工学系
〒441-8580 豊橋市天伯町字雲雀ヶ丘1-1

近年、音声対話システムにおける音声認識技術や言語処理技術の要素技術に関しては確立しつつあり、実用化に向けシステム開発が進んでいる。実用化などを考慮した場合、今までのような使い易さ、頑健性などに関する技術だけでは不十分であり、拡張性や移植性なども十分考慮する必要がある。
一方、音声認識時に使用する言語モデルを構築するには多くの学習文が必要である。しかし、あるタスク／ドメインに限定した言語モデルを構築しようとしても、数千文もの学習文を用意するのは困難である。そこで、数百文程度の文集合からでも、登録語に対してはそのクラス名と読みを登録するだけで性能の良い言語モデルを構築できる方法を提案する。本手法では、既知語には単語バイグラムを、未知語にはクラスバイグラムを利用することで精度の良い言語モデルを構築した。

Language Model Construction based on Unknown Word Registration for Spoken Dialogue Systems

Satoru KOGURE, Kenji Hori and Seiichi NAKAGAWA

Department of Information and Computer Sciences
Toyohashi University of Technology, Tenpaku, Toyohashi, 441-8580, Japan
{kogure, ken, nakagawa}@slp.tutics.tut.ac.jp

Recently the technology for speech recognition and language processing for spoken dialogue systems has been improved, and speech recognition systems and dialogue systems have been developed to be practical use. But in order to become practical, not only those fundamental techniques but also the techniques of portability and expansibility should be developed.
We need many user utterances to construct a language model for speech recognition. However it is difficult to construct the language model for a new task/domain because many utterances cannot be collected easily. So, we propose a method that is able to construct an efficient language model, preparing at least only several hundreds utterances. In the method, the language model is constructed by the word bigram and class bigram for known words and unknown words, respectively.

1 はじめに

近年、音声対話システムの研究が広く行なわれている。最近では特に「ロバスト性」や「ユーザビリティー」に関する研究が、注目を集めている。しかし、これらのシステムはあるタスクやドメインに限定した上での開発であり、処理データにはタスクやドメインに依存している部分がかなりあると考えられる。

現在、音声認識システムはかなり実用化に近いものが開発され、音声対話システムもこれから益々実用になると思われる。しかしながら、実際に新しい

音声対話システムを構築するには莫大なコストがかかり、今後、これまで開発してきたシステムを他のドメイン用に変更したり、汎用的なシステムを開発することが重要になってくることは間違いない。実際、「移植性」や「拡張性」を重要視する研究も行なわれてきている [1][2][3][4][5][6][7][8][9]。

PIA[1] というシステムは、複雑な音声対話システムでもプロトタイプを簡単に構築できる。Visual BASIC を用いて実装され、認識のロバスト性と対話の自然さを両立させることに重点をおいている。REWARD(Real World Applications of Robust Dialogue)[2] というプロジェクトで試作しているシステムは、開発者がシステムの開発、デバッグを一括して管理でき、従来の音声対話構築よりも早い時間でシステムを構築することが出来る。OGI (Oregon Graduate Institute) で研究が勧められている CSLU Toolket[3] というシステムは、音声に関する知識を全然持っていなくても音声を使ったアプリケーションを素早く構築することが出来る。笹島らは EUROPA という音声対話構築ルールを提案し、MINOS というカーナビゲーションシステムを実際に構築して評価している [4]。田中らは、「情報検索」にタスクを限定することで、データベース自体には依存しない汎用的な情報用検索音声対話についての研究を報告している [5]。このシステムでは、アプリケーション開発者は、複雑な文法、語彙の設定を行なうことなく、手持ちのデータベースに対しての音声対話による情報検索を利用することが出来るが、ユーザの発話できる内容をメニューに提示することでユーザの発話を制限している。 秋葉らは、マルチモーダルインターフェースの汎用性に関する研究を報告している [6]。彼らは、MILES というマルチモーダル対話記述言語を開発し、ジャンケンや地下鉄乗り換え案内などいくつかの対話タスクを実際に試作している。 荒木らは、音声対話システムにおけるタスクを「スロットフィリング」「データベース検索」「説明」「それ以外」に分割している [7]。前3つのタスクについては、その対話をスクリプト言語で記述できる。我々も「音声対話システムにおける移植性」に関する考察を行なっている [8][9]。

一方、言語モデルは、一般的に認識対象となるタスク毎に構築することが望ましい。ある新しいタスクにおけるユーザ発話文を認識するためのバイグラムを構築するためには、該当タスクの発話文集合が数千文は必要となることが分かっている [10]。しかし、タスクが変更されるごとに収集を繰り返すことは非常にコストを要する。最近、少量のコーパスを使って言語モデルを適応したり、単語バイグラムとクラスバイグラムを統合して使用する手法が提案されている [11][12][13][14][15][16][17]。

伊藤らは大量の一般テキストと少量の特定タスクテキストを混合して言語モデルを構築する手法を提案している [11]。小林紀彦らは単語バイグラムとクラスバイグラムの両者を求めておき、出現頻度の低い単語についてはクラスバイグラムを使用し、出現頻度の高い単語については単語バイグラムを用いる手法 [12] を提案し、通常のバイグラムよりも認識率が良くなったことを示している。小林彰夫らはタスクごとに別に言語モデルを構築し認識する際に確定した1単語ごとに、それぞれの言語モデルの重みを計算して重みつき和を実際の言語モデル確率として採用する手法を提案している [13]。佐々木らは単語間の関連性を用いてバイグラム確率を推定する手法を提案している [14]。この手法では、大量の不特定分野のテキストコーパスと少量の目的タスクのテキストコーパスから、両者に現れるタスク独立語を求め、このタスク独立語ではない、タスクに関連した単語の関連性のみを考慮している。大附らは大量のタスク外のテキストコーパスから、対象タスクに近いデータを抽出することでタスク適応を行なう手法を提案している [15]。小笠原らは大量テキストに対して適応テキストに N倍の重みをつけて混合したテキストを用いてバイグラムを構築している [16]。この手法は単語バイグラムとクラスバイグラムの両者の性質を持っており、閾値を動かすことで、単語バイグラムに近付けたりクラスバイグラムに近付けることが可能である。これらの研究は文献 [11] を除いていずれも新聞記事やニュース音声の言語モデルの適応化について考察されたものである。

本研究の目的は、ある特定のタスクの音声対話システムで使用するために、未登録後 (未知語) に対してそのクラス名と読みを登録することで精度の良い言語モデルを構築する手法を開発することである。該当タスクでの比較的小さな、200 文程度の文集合と、データベース検索に関連した様々なタスクでの文集合からバイグラムを構築する二つの方法を比較した。一つは固有名詞に関しては1つのクラス、又は複数のクラスとしてバイグラムを求める方法である。もう一つは、データベース検索に関連した対話文集合からクラスバイグラムを求め、該当タスクの文集合から単語バイグラムを求め、この二つのバイグラムを用いて、クラス名を人手で付与した未知語のバイグラム確率を推定する方法である。

2 通常の言語モデルによる認識

未登録語 (未知語) を考慮した言語モデルとの比較を行なうために、まず、通常の言語モデルでの評価結果を示す。

2.1 言語モデルの構築

文献検索をタスクとする文集合 [18] と、データベース検索に関連した対話文集合 [19][20] を使用して評価を行なった。

- 文献検索対話文集合 (以下、文献対話文) 495 文
- データベース検索や案内に関連した対話文集合 (道案内、地図課題タスク、秘書システム、スケジュール管理、クロスワードパズル、地理・旅行案内、レテフォンショッピング、スケジュール調整タスク、留学生の夏休み、間違い探し、スキープラン対話：以下、関連対話文) 8927 文

まず、テスト用に文献対話文 495 文から、100 文 (818 単語) を抽出した (テスト文献対話文)。学習には文献対話文 395 文と、関連対話文 8927 文を使用した。表 1に生成した bigram のパープレキシティとカバレージを示す。PP はパープレキシティを、APP は補正パープレキシティを示し、APP を算出する際の未知語の確率は 1/50000 としている。mix 5 とは、関連対話文 8,927 文に対する文献対話文 200 文の相対頻度を 5 倍したものを示す。具体的には、関連対話文 8,927 文 と 文献対話文 1,000 文 (同じ文を 5 回ずつ) を学習に使用したことになる。同様に mix 10 は、相対頻度を 10 倍したものである。また、比較のために、約 47 万文のニュース文から語彙数 20,000 の言語モデルを構築した。さらに、市販の音声ディクテーションソフトウェア A, B も評価した。文献対話文 200 文 と、関連対話文と文献対話を混合させた mix 5, mix 10 とを比較すると、カバレージが 89.5% から 93.9% と改善されたが、逆にパープレキシティは高くなった。

表 1: bigram のパープレキシティとカバレージ

言語モデル	語彙数	パープレキシティ		Cov.[%]
		PP	APP	unigram
文献対話文 100 文	259	22	83	82.9
文献対話文 200 文	373	20	45	89.5
文献対話文 395 文	569	18	29	94.0
関連対話文 8927 文	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	20000	880	1766	86.5
市販ソフト A	約 40000	—	—	—
市販ソフト B	約 80000	—	—	—

2.2 音声認識実験

男性 3 名からテスト文献対話文 100 文の音声を収録し、前節の各言語モデルを用いて、音声認識実験を行なった。テスト文は、「えと大野先生という人の論文を探しているのですが」「キーワードは解析学、出版年は 1980 年でお願いします」のような文集合である。音響モデル、特徴パラメータ等は表2に示す通りである。単語認識率を表3に示す。Acc. は単語正解精度、Cor. は単語正解率を示す。文献対話文 395 文 が最も認識率が良かった。また、文献対話文 200 文 のモデルと、mix 10 のモデルの認識率はほぼ同じであった。また、市販のディクテーションソフトは文献対話文 100 文による言語モデルの場合よりもやや悪い認識率であった。

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)+ΔPOW+ΔΔPOW

表 3: 正解精度と正解率

言語モデル	Cor.(%)	Acc.(%)
文献対話文 100 文	57.7	46.0
文献対話文 200 文	65.4	57.0
文献対話文 395 文	71.7	65.9
関連対話文 8927 文	49.0	42.1
mix 5	60.5	53.2
mix 10	64.9	56.9
ニュース文	30.0	21.2
市販ソフト A	54.2	45.1
市販ソフト B	51.3	43.3

3 固有名詞の登録

テスト文に対する各言語モデルの未知語総数、未知語の種類数、未知語中の固有名詞の数を表 4に示す。これを見ると、言語モデル mix の未知語総数が文献対話文 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対話文 200 文の場合と変わっていないことから、関連タスク 8927 文では、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、文献対話文 200 文から作成した言語モデルの語彙に未知語の固有名詞を追加しそのユニグラム確率を 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。固有名詞のクラスは、 人名、団体名、国名、国外都市、国外地域、国内大地域、国内小地域、海、山、書名、店名の 11 クラスである。表 5に結果を示す。固有名詞を登録することにより単語正解精度は 8〜9%上昇し、文献対話文 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が結果に大きく影響することがわかる。

表 4: 各言語モデルでのテスト文 100 文 (818 単語) 中の未知語数

言語モデル	未知語		
	総数	種類数	固有名詞
文献対話文 100 文	147	120	31
文献対話文 200 文	101	90	27
文献対話文 395 文	58	53	14
関連対話文 8927 文	160	91	43
mix 5 倍	60	51	27
mix 10 倍	60	51	27
ニュース文	128	78	38

表 5: 文献対話文 200 文から求めた言語モデルに未知語の固文有名詞を登録した時の単語正解精度と単語正解率

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

4 クラスバイグラムの構築

前節で、固有名詞のみをクラスバイグラム化した際の評価結果を示した。本節では、固有名詞だけでなく、全単語の言語生起確率にクラスバイグラムを用いた場合の評価結果を示す。

クラスバイグラムを用いると、単語バイグラムは一般的に式 (1) で計算できる。

$$P(w_2|w_1) = P(c_2|c_1)P(w_2|c_2) \quad (1)$$

ここで、c_i は単語 w_i のクラス名を示す。また、$P(c_2|c_1)$ はクラスバイグラムであり、$P(w_2|c_2)$ はクラス c_2 から 単語 w_2 を生成する確率である。クラスは、基本的に JUMAN3.2 での品詞を用いる。例外として、助詞を「は」「が」「の」「を」「に」「と」「それ以外」に、固有名詞を前述の 11 クラスに、普通名詞を図 1に示すアルゴリズム [9] で得られた 100 クラスに 分割した。この結果、総クラス数は、133 になった。また、クラスから単語への生起確率 $P(w|c)$ は学習に使用しコーパスでの出現頻度を考慮して定義した。

EDR の概念辞書に含まれる 20 万概念を 100 概念にクラスタリングする。

1. EDR の日本語コーパスから 概念毎の頻度 $freq(i)$ を求めておく。(i は概念子)
2. EDR の概念関係辞書から、上位概念集合 $up\_context(i)$ と、下位概念集合 $down\_context(i)$ を求める。
3. 下位概念集合 $down\_context(i)$ が空である 概念 i のうち、$freq(i)$ が一番小さい概念 i_d について、
   (a) 上位概念集合 $up\_context(i_d)$ の要素 j について、下位概念集合 $down\_context(j)$ から、概念子 i_d の枝を除去し、概念子 j とマージする。
4. 3 の操作を、総概念数 100 になるまで繰り返す。

図 1: 普通名詞のクラスタリング

まず、これまでに述べた方法で構築した言語モデルのパープレキシティとユニグラムのカバレージを表 6に示す。表中の括弧で示されている値は、全未知語を語彙に登録した時のパープレキシティとユニグラムのカバレージである。表 1と比較すると、クラスバイグラムにすると、カバレージが上昇するが、パープレキシティも高くなってしまうことが分かる。

認識結果を表 7に示す。表中の括弧内の値は、未知語を語彙に登録した時の正解精度、正解率を示している。クラスバイグラムを用いない表 3と比較すると全般に認識率は向上しているが、固有名詞のみをクラスバイグラムで構築した表 5 の結果と比較すると認識率が低くなった。また、未知語を語彙に登録すると更に認識率が向上することが分かった。

表 6: クラスバイグラムによるテスト文パープレキシティとカバレージ (括弧内は未登録後を追加した場合の値)

言語モデル	PP	APP	Cov.[%]
文献対話文 200 文	61(79)	323(79)	89.5(100)
関連対話文 8927 文	142(137)	971(137)	81.0(100)
mix 5	91(99)	666(99)	93.9(100)
mix 10	88(93)	653(93)	93.9(100)

表 7: クラスバイグラムによる単語正解精度と単語正解率 (括弧内は未登録後を追加した場合の値)

言語モデル	Cor.[%]	Acc.[%]
文献対話文 200 文	65.2(70.6)	56.4(66.9)
関連対話文 8927 文	60.1(66.0)	50.3(61.1)
mix 5	63.4(69.6)	55.0(66.0)
mix 10	64.3(70.2)	55.9(66.7)

5 未知語を考慮した単語バイグラムとクラスバイグラムの利用

これまでに、全単語をクラスバイグラム化するよりも固有名詞のみをクラスバイグラム化すると言語モデルの性能が良くなること、及び、未知語を語彙に登録することで言語モデルの性能が良くなることが分かった。よって、本節では、学習文中に含まれる単語については単語バイグラムを、未知語についてはクラスバイグラムを利用する方法について考察する。

ある学習コーパス (語彙集合を Vとする) から既に単語バイグラム $P'(w_2|w_1)$ と、クラスバイグラム $P'(c_2|c_1)$ が求まっているとする。あるテスト文における bigram の言語確率を $P(w_2|w_1)$ とする。しかし、この式は、$w_1, w_2 \in V$ の時にしか使えない。よって、(1) w_1, w_2 ともに既知語、(2) w_2 が既知語、w_1 が未知語、(3) w_2 が未知語、w_1 が既知語、(4) w_1, w_2 ともに未知語の 4 通りについてそれぞれのバイグラムを計算する必要がある。以下に算出式を示す。(3) における $C(w_2, c_1)$ は 単語 w_2 とクラス c_1 に属する各単語のペアの出現回数つまりは、$\sum_{w_i \in c_1} C(w_2, w_i)$ で求める。

$$
\begin{aligned}
(1)\quad P(w_2|w_1) &= P'(w_2|w_1) \\
(2)\quad P(w_2|w_1) &= P(c_2|w_1) \times P(w_2|c_2) \\
&= \sum_{w_i \in c_2} P'(w_i|w_1) \times P(w_2|c_2) \\
(3)\quad P(w_2|w_1) &= \frac{C(w_2, c_1)}{C(c_1)} \\
(4)\quad P(w_2|w_1) &= P'(c_2|c_1) \times P(w_2|c_2)
\end{aligned}
$$

この言語モデルを文献文集合 200 文の学習文から構築しパープレキシティと音声認識率を求めた。クラスについては前節で述べたクラス分類を使用した。結果を表 8と表 9に示す。表 8の Cov. はユニグラムのカバレージを示している。表 9 を見ると、文献対話文 395 文で学習した単語バイグラムよりも、文献対話文 200 文から単語バイグラムとクラスバイグラムを求め、上に示した計算式で構築し直したバイグラムの方が、学習文数が少ないにも関わらず、単語正解精度、単語正解率ともに向上した。これにより、本手法の有効性が示されたと考えられる。

表 8: 単語バイグラムとクラスバイグラムの利用によるテスト文パープレキシティと カバレージ

言語モデル	PP	APP	Cov.[%]
文献対話文 200 文	21	21	100.0

表 9: 単語バイグラムとクラスバイグラムの利用による単語正解精度、単語正解率 (表 3 参照)

言語モデル	bigram	Cor.[%]	Acc.[%]
文献対話文 200 文	単語	65.4	57.0
文献対話文 200 文	単語, クラス	75.1	72.7
文献対話文 395 文	単語	71.7	65.9

6 まとめ

音声認識時に避けて通れない未知語に対して様々な検討を行なった。まず、未知語であった固有名詞を語彙に登録することによって、認識率が向上した。また、固有名詞を 11 個のクラスにクラスタリングすると、さらに認識率が向上した。次に、全単語を

133 クラスにクラスタリングし、クラスバイグラムを使用した所、認識率が低下した。これらの結果から、既知語は単語バイグラムで、未知語部分のみをクラスバイグラムで求めると言う手法を提案した。文献対話文 200 文をこの手法で学習した言語モデルは、文献対話文 395 文の単語バイグラムと比較して、文数が約半文であるにも関わらず、認識率が良くなると言う結果を得た。これは、本研究で提案した、少量の学習コーパスから、単語バイグラムとクラスバイグラムを求めておき、未知語のクラスを使って言語確率を求める手法の有効性を示していると考えられる。

現在、このクラスタリングは半自動であり、まだ手作業の部分が存在する。実際の所、未知語を完全に自動で解析することは不可能であり、手作業でのクラスの付与は必要なことであると考えられる。今後は、任意の単語を対話的にクラスタリングできる手法を考案する必要がある。また、音声対話システムにおける対話状況に応じた bigram 辞書の切替えや関連対話文集合の効果的利用法の検討なども行なっていきたい。

参考文献

[1] Kaspar.S, Hoffmannn.A: "Semi-automated incremental prototyping of spoken dialog systems" ,Proc. ICSLP 98, Vol.3, pp.859–862(1998)

[2] Brondsted.T, Bai.B, Olsen.J: "The REWARD Service Creation Environment. An Overview", Proc. ICSLP 98, Vol.4,pp.1175–1178(1998)

[3] Stephen Sutton, Ronald Cole, Jacques de Villiers, Johan Schalwyk, Pieter Vermeulen, Mike Macon, Yonghohg Yan, Ed Kaiser, Brian Rundle, Khaldoun Shobaki, Paul Hosom, Alex Kain, Johan Wouters, Dominic Massaro, Michael Cohen: "Universal Speech Tools:The CSLU Toolkit", Fr2C4, Proc. ICSLP98, Vol.7 pp.3221-3224

[4] M. Sasajima, T. Yano, and Y. Kono: "Europa: Generic Framework for Developing Spoken Dialogue System", Proc. EUROSPEECH99, pp.1163–1166.

[5] 田中 克明, 川原 達也, 堂下 修司: 「汎用的な情報検索音声対話プラットホーム」, 情報処理学会, 情処研報, SLP-24-14, 1998

[6] 秋葉 友良, 伊藤 克亘: 「スクリプト言語を用いたマルチモーダル対話記述の試み」, 情報処理学会, 情処研報, SLP-23-1/HI-80-1, 1998.10

[7] 荒木 雅弘, 駒谷 和範, 平田 大志, 堂下 修司: 「音声対話システム構築のための対話ライブラリ」, 人工知能学会, SIG-SLUD-9901-1, 1999.1

[8] 小暮 悟, 伊藤 敏彦, 中川 聖一: 「音声対話システムの移植性に関する考察 –観光案内システムとデータベース検索システム–」, 情報処理学会, 情処研報, SLP-25-3, 1999.2

[9] 小暮 悟, 中川 聖一:「移植性の高いデータベース検索用音声対話システムの試作」, 情報処理学会, 情処研報, SLP-27-15, 1999.7

[10] 中川 聖一, 大谷 耕嗣: 「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)

[11] 伊藤 彰則, 好田 正紀: 「対話音声認識のための事前タスク適応の検討」, 情報処理学会, 情処研報, SLP-14-13 (1998.3)

[12] 小林 紀彦, 小林 哲則:「クラス統計と単語統計の併用による小規模学習データのための統計的言語モデル構成法」, 自然言語処理学会, NL-131-1/SLP-26-1 (1999.5)

[13] 小林 彰夫, 今井 亨, 田中 英輝, 安藤 彰男: 「ニュース音声認識のための言語モデルの動的適応化」, 日本音響学会 春季研究発表会 講演論文集,3-8-2(2000.3)

[14] 佐々木 耕樹, 広瀬 啓吉:「単語間の関連性を用いた言語モデルのタスク適応」, 日本音響学会 春季研究発表会 講演論文集,3-8-3(2000.3)

[15] 大附 克年, 堀 貴明, 川端 豪, 小川 通朗, 北脇 信彦:「大規模データベースを用いたタスク依存言語モデル構築の検討」, 日本音響学会 春季研究発表会 講演論文集,3-8-4(2000.3)

[16] 小笠原 教充, 加藤 正治, 伊藤 彰則, 好田 正紀:「品詞と高頻度単語の N-gram を使用したタスク適応の検討」, 日本音響学会 春季研究発表会 講演論文集, 3-8-5(2000.3)

[17] 堀 賢史, 小暮 悟, 中川 聖一:「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」, 情報処理学会, 第 60 回全国大会, 1ZA-1, 2 分冊, pp.259–260(2000.3)

[18] 藤崎 博也, 大野 澄雄, 飯島 岐勇:「音声対話の収録方法とその設定について」, 学振未来開拓事業プロジェクト 音声対話資料収録用参考資料, 1998.1

[19] 文部省重点領域研究 [音声対話],CD-ROM Vol.1(1994),Vol.2-4(1995)

[20] 電総研道案内対話音声コーパス,CD-ROM Vol.1-7(1998)