

TR-IT-0135

未知語処理のための音響モデルの検討
On the study of garbage cluster model
for unknown word detection

若杉 和徳 坂本 博之
Kazunori Wakasugi Hiroyuki Sakamoto
 松永 昭一
 Shoichi Matsunaga

1995.9

実際の連続音声認識システムでは、単語間の連結共起に関する意味的な制約を構文規則の設計に反映させ、単語の予測能力を高めている。しかし、入力が辞書に登録されていない単語（未知語）を含む場合は構文規則で受理されず認識精度が低下する。これを防ぐためには未知語処理が必要となる。今回の実習では未知語検出の方法として Garbage HMM の利用を検討した。

©ATR 音声翻訳通信研究所

©ATR Interpreting Telecommunications Research Laboratories

1 はじめに

音声認識技術の進歩と共に音声認識研究における認識対象は、数字や連続数字に代表される孤立単語や連続単語から大語彙の連続音声認識へと広がってきた。この連続音声認識は、孤立単語認識などに比べ音響的特徴パターンの種類や変化が多いので高い精度での認識が困難である。そのため実際の連続音声認識システムでは、単語間の連結共起に関する意味的な制約を構文規則の設計に反映させ、単語の予測能力を高めてきた。簡単な構文規則の例を図1に示す。

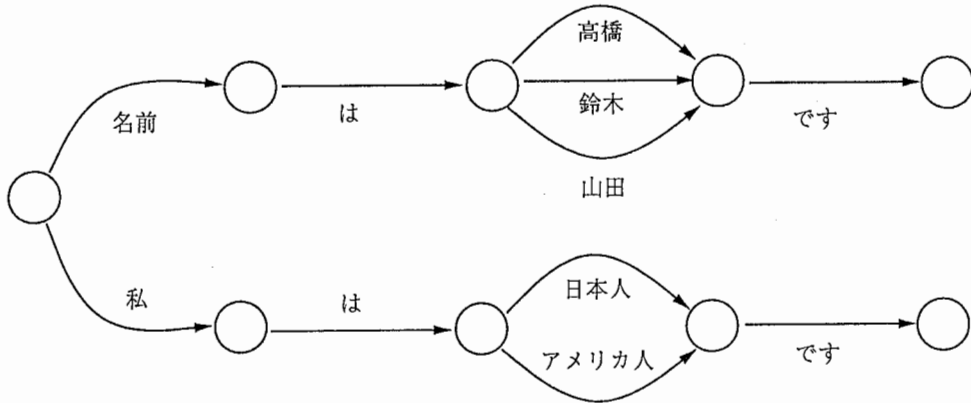


図 1: 簡単な構文規則の例

しかし、さらに認識対象が制限の多い朗読調の連続音声から、日常会話に近い発話である「自由発話」に広がってくると、この制約を満たさない入力 that 想定される。例えば図1の構文規則では「名前は高橋です」、「私は日本人です」といった発話は正しく受理されるが、「私は後藤です」という発話は構文規則上に「後藤」という名詞がないので、パスが存在せず正しく受理されない。また「名前は小林です」、「私はロシア人です」といった発話なども「小林」、「ロシア人」などの辞書に登録されていない単語(未知語)は構文規則に含まれていないので受理されない。このために未知語は認識精度の低下の原因となる。これを防ぐために未知語を含んだ発話に対して精度良く認識できる処理が必要と考えられる。

2 未知語処理のための音響モデルの検討

実習の目的は、未知語処理を行なう上でどのような音響モデルを利用すれば少ない処理量で高い未知語検出率が得られるかを検討することである。今回は未知語検出の方法として Garbage HMM の利用を検討してみる。

Garbage HMM とは数種類のキーワードを含む自由発話から、キーワードスポッティングを行なう方法として用いられるキーワード以外の任意の音声、非音声のモデルのことである。今回の実習ではこの考え方を拡張し、いくつかの音素を一つの音響的モデルを用いてモデル化することで、自由発話文の認識における未知語処理に応用することを検討する。

Garbage HMM を使うために図 1 の構文規則を図 ?? のようにして用いる。

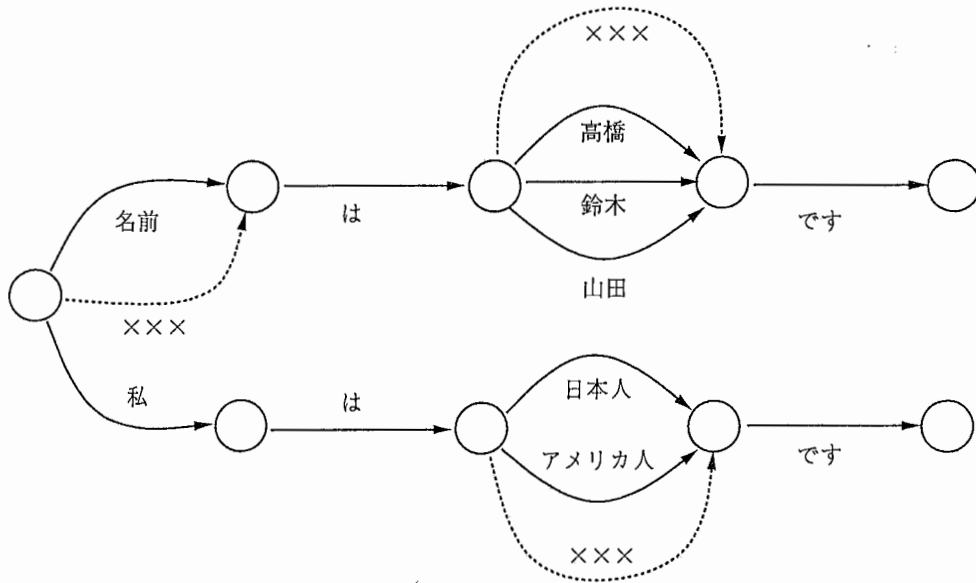


図 2: Garbage HMM を使った構文規則の例

3 Garbage HMM の作成

Garbage HMM には、音響的に類似している音素をまとめたモデルを考える。そこで SSS (Successive State Splitting algorithm) によって自動的に分割された allophone を調べ音素のクラスタを決めた。これらの SSS により分けられた音素を表 1 に示す。

表 1: SSS により分けられた音素

クラスタ数	分けられた音素
2	-(a i u e o n g q j k zh z d m g ch r sh ts s b t w n p h)
3	-(a o w)(i u e n g q j k zh z d m g ch r sh ts s b t n p h)
4	-(a o w)(e q j k zh z d m g ch r sh ts s b t n p h)(i u n g)
9	- o w a e (i u n g) j (m g n) (q k zh z d ch r sh ts s b t p h)
12	- o w a g (ch sh ts s) (d r b) (q k zh z t p h) e (i u n g) j (m n)
14	- o a w g (ch sh) (d r b) (q k zh z t p h) (ts s) (u n g) (m n) e i j
20	- a w o r (k t) q (u n g) (m n) e i j g (d b) z p (ts s) (ch sh) zh h

今回の実習ではどのような音素のクラスタから作成した Garbage HMM が未知語検出に適しているかを調べるのが目的である。今回は 2,3,4,9,12,14,20 個に音素のクラスタを分けた Garbage HMM を作成した。この際まとめた一つの音素当たり混合数 5 としてクラスタ内の音素の分だけ Garbage HMM の混合数を増やした。

例 1) a, o, w の 3 つの音素をまとめた Garbage HMM の混合数は 15

例 2) q, k, zh, z, t, p, h の 7 つの音素をまとめた Garbage HMM の混合数は 35

また学習データは男性 1 人 (MAU) による 2620 単語を使用した。

4 Garbage HMM を使った未知語を含む文の認識実験

未知語検出のための Garbage HMM 利用の有効性を検討するために以下のような実験を行った。

4.1 実験条件

実験には SSS-Toolkit(Ver3.0) を使用した。使用音素モデルの仕様は表 2 に示す。

表 2: 音素モデルの仕様

サンプリング周波数	12 [kHz]
分析窓	20 [ms] ハミング窓
フレームシフト	5 [ms]
パラメータ	16 次 LPC ケプストラム + log パワー + 16 次 Δ ケプストラム + Δ log パワー
音素モデル	26 個のコンテキスト非依存 HMM (4 状態 5 混合)
学習データ	男性 1 名 (MAU) (2620 単語)
使用音素ラベル	w,k,u,e,p,g,a,j,n,ts,ch,r,i, o,q,t,sh,N,zh,m,d,b,z,s,h,-(無音)
ビーム幅	3000

4.2 Garbage HMM を使用した認識実験

音響モデルにはコンテキスト非依存の音素 HMM を使い、未知語区間で Garbage HMM を使用する。また文法には国際会議予約 (語彙数:454)、評価データには学習データと同じ話者 (MAU) による 57 文を用いた。

実際には文章中のいろいろな場所に未知語が含まれる可能性があるが、今回の実験では未知語の現れる場所を名詞だけに限定して実験を行なった。

認識結果を表 3 に示す。但し表中の total は評価データ (男性 1 名 (MAU) による 57 文) 中の全語句数、I は挿入誤り、D は消去誤り、S は置換誤りを表す。認識精度は以下の式により求めた。

$$\text{認識精度} = \frac{M - I - D - S}{M} \times 100 (\%)$$

表 3: Garbage HMM を使用した認識実験結果

使用 Garbage HMM の数	total	I	D	S	認識精度 [%]
2	363	10	170	36	40.5
3	363	6	162	26	46.6
4	363	8	127	23	56.5
9	363	9	176	31	40.5
12	363	7	188	23	39.9
14	363	7	183	29	39.7
20	363	5	205	17	37.5

* 未知語処理なしの場合の認識精度は 59.0[%] だった。

結果は使用 Garbage HMM の数が4つのときの認識精度が最も高い。誤りの種類を見てみると、どの Garbage HMM の数でも消去誤りが一番多い。これは認識するときに名詞だけをうまく検出できずに前後の形容詞などもいっしょにひとつの名詞と認識してしまったためである。例えば「名前は鈴木まゆみです。」という文章では名詞である「名前」と「鈴木まゆみ」のふたつを未知語として検出しなくてはならないが、誤って「名前は鈴木まゆみ」をひとつの未知語として検出してしまうといったことである。このことから GarbageHMM のスコアは構文規則上のものより高くなり易いことがわかる。

4.3 Garbage HMM および N-gram を使用した認識実験

次に Garbage HMM に N-gram モデルを使用してその効果を調べるための認識実験を行なった。N-gram モデルとは使用ラベルの N 組の連鎖の確率をモデル化したものである。この N-gram モデルの学習には国際会議予約タスクの対話テキストデータから名詞のみ(全名詞数 58896; 異なり名詞数 5072) 集めたものを、それぞれの Garbage HMM で使用したラベルに変換して使用した。

ここでは 4-gram モデルを使用し、次の式に従い認識候補の未知語区間のスコアを求めた。

$$P_{total} = \log(P_{hmm}) + \omega * \log(P_{gram})$$

ここで P_{total} は認識候補の未知語区間のスコア、 P_{hmm} は Garbage HMM と入力を照合してえられるスコア、 P_{gram} は Garbage HMM の 4-gram のスコアである。 ω は 4-gram モデルの重みで今回は 1.0 とした。

この 4-gram モデルを使用した場合の認識実験結果を表 4 に示す。

表 4: 4-gram モデルを使用した場合の認識結果

使用 Garbage HMM の数	total	I	D	S	認識精度 [%]
2	363	6	173	35	41.0
4	363	10	62	31	71.6
9	363	5	108	35	59.2
14	363	6	93	38	62.3
20	363	9	94	39	60.9

4-gram モデルを使用した場合、使用しない場合より認識精度が改善された。その中で認識精度が最もよかったのは、使用 Garbage HMM の数が4つのときである。しかし 4-gram モデルを使用して認識精度が向上した割合は、使用 Garbage HMM の数が 9,14,20 のときの方が高い。今回の実験では 4-gram モデルの重みがすべて 1.0 としてあるが、それぞれの Garbage HMM によりその最適重みは違うと考えられる。従ってそれぞれの最適重みを求めてから比較を行なった方がよいが今回は時間の都合によりできなかった。

4-gram モデルを使用した場合の誤りの種類を見てみると、使用しない場合より消去誤りが減ったが、逆に置換誤りが増えている。これは本来未知語(この実験では名詞)のある場所に誤って他の語を認識してしまったためである。例えば「期限は今年いっぱいです」という文章があったら、「期限」と「今年」という名詞を検出して「(未知語)は(未知語)いっぱいです」と認識

しなくてはならないのに、「決まって欲しい(未知語)です」という全然違う結果になってしま
うというようなことである。原因は4-gram モデルを使用しない場合とは逆に構文規則の方のスコアが高くなってしまったことと考えられる。Garbage HMM の数が多い程この傾向がある。

5 Garbage HMM を使用した場合の処理速度の比較

さらに Garbage HMM のモデル数の違いによる処理量を比較するために、それぞれの場合の認識処理時間を調べた。この結果を以下に示す。この処理時間は評価データ中の34文の認識処理時間の合計である。認識処理時間は4-gram モデルなしの場合は Garbage HMM の数が4つ

表 5: 認識処理時間

Garbage hmm のモデル数	N-gram モデルなし [sec]	N-gram モデルあり [sec]
2	3518	3931
4	3127	3069
9	4951	3868
14	7692	4975
20	12315	6581

のときが最も少ない。あとは Garbage HMM の増加と共に認識処理時間が多くなっている。例外は Garbage HMM が2つのときで4つのときより多くなっているが、これはスコアの高い認識候補が多くできてしまったためである。4-gram モデルありの場合も認識処理時間の少なさの順位は4-gram モデルなしの場合と大差ない。また4-gram モデルありの方がなしの場合より認識処理時間が少なくなっている。これは4-gram モデルによる認識候補の枝刈りがうまくいったためと考えられる。例外は Garbage HMM が2つのときで、これは4-gram モデルによって逆に認識候補が増えてしまったためと考えられる。

6 考察

今回の実験結果だけからどのような Garbage HMM が未知語検出に適しているかを判断するのはサンプルが少なく、また N-gram モデルの重みによって認識精度は変わってくるので難しいと考えられる。そこで今後の課題としてより多くの種類の Garbage HMM の認識精度の比較、さらにそれぞれの Garbage HMM の最適重みにより N-gram モデルを使った認識精度の比較が必要と考えられる。

7 謝辞

快適な実習環境を提供して頂いた ATR 音声翻訳通信研究所第一研究室の연구원諸氏に感謝の意を表します。また、本実習の機会を与えて下さった ATR 音声翻訳通信研究所の山崎泰弘社長、並びに ATR 音声翻訳通信研究所第一研究室の匂坂芳典室長に深く感謝致します。

参考文献

- [1] 伊藤克亘, 速水悟, 田中穂積.” 連続音声認識における未知語の扱い”, 信学技報 SP91-96, pp.41-47, 1991.12
- [2] 小沼知浩, 武田一哉.” 自由発話文理解のための Garbage HMM の利用の検討”, 信学技報 SP92-127, (1993-01)
- [3] 伊藤克亘, 速水悟, 田中穂積.” 未知語処理のための言語的制約について”, 音響学会講演論文集, pp.49-50, (1993)
- [4] 伊藤克亘, 速水悟, 田中穂積.” 音声対話システムにおける未知語の扱い”, 人工知能学会研究会資料, pp.1-9, 1992.4
- [5] 井上直巳, 武田一哉, 山本誠一.” ガーベジ HMM を用いた自由発話文中の不要語処理手法”, 電子情報通信学会論文誌 Vol.J77-A No.2 pp.215-222, 1994.2