

## タスクに無関係な発話を聞き流すための言語モデル構築法\*

高岡充<sup>†</sup> 上平拓弥<sup>††</sup> 西崎博光<sup>†††</sup> 関口芳廣<sup>†††</sup>

<sup>†</sup>山梨大学工学部 <sup>††</sup>大学院医学工学総合教育部 <sup>†††</sup>医学工学総合研究部

### 1 はじめに

筆者らは複数人と対話可能な音声対話ロボット『キクラゲ』[1]を開発している。音声対話ロボットの音声認識機能を実現するためには専用の言語モデルが必要である。これを実現するため、タスクに関係した対話を大量に録音し、書き起こし文を用意することが一般的である。しかし、音声の書き起こし作業は手間とコストがかかるため、対話ロボット開発の障害になる場合も多い。

そこで、タスクに出現するキーワードをクラス分けしたクラス言語モデルを構築する手法をとっている。これにより少量の学習テキストで言語モデルを構築することができるが、この言語モデルはキーワードのみから構成されるため、タスクと無関係な発話でキーワードの湧き出し誤りが起こり、システムの誤動作が多発する。

本研究ではクラス言語モデルにガベージクラス [1]を導入し、キーワードの湧き出しが少ない言語モデルを構築する手法を提案する。実際にロボットの言語モデルにクラス言語モデルとガベージクラスを導入したところ、タスクに無関係な発話が存在しても、89%の確率で正しい音声認識が可能になった。

### 2 音声対話ロボット

#### 2.1 タスクの概要

音声対話ロボットのタスクとして、カードゲームを行うことを想定している。このカードゲームはユーザーとロボットが対話を行い、バツ抜き要領でカードをやり取りするものである。

対話処理には音声認識が必須であり、発話の意図を理解する上で重要な単語（キーワード）を認識することで、ロボットはカードゲームを進行することができる。

#### 2.2 言語モデルの構築方法

このロボットでは言語モデルの学習を効率的に行うために、クラス言語モデルを利用している。これは意味を考慮してキーワードを事前にクラスに分類し、クラス単位で N-gram の学習を行うものである。キーワードをクラスに分類したものを表 1 に示す。

キーワードは計 108 個存在しており、これらをクラス分けした計 12 種のクラス名で学習文章を作成する (図 1)。このようにすることで、対話に必要なキーワードを効率的に学習することが可能になる。

表 1 キーワードのクラス分類

クラス名	登録数	例
所有	9	持ってるよ
順番	5	の番だよ
カード名	20	カエルのカード
要求	9	ちょうだい
勝利	4	勝ったよ
あいさつ	4	ありがとう
人名	14	〇〇くん
不所持	11	持ってないよ
否定	15	ダメ
催促	2	ロボの番だよ
再要求	8	何だって
肯定	9	良いよ

クラス単位の学習	学習できる文章例
<人名><カード名><要求>	〇〇くんカエルのカードちょうだい △△さんクジラのカードください
<人名><順番>	◇◇くんの番だよ □□さんの番
<あいさつ>	おはよう ありがとう

図 1 言語モデルの学習文章例

#### 2.3 クラス言語モデルの問題点

キーワードをクラス化した学習テキストだけでは、カードゲームの発話と無関係な発話（以下、ドメイン外発話）においてキーワードの湧き出し誤りが多発する。湧き出しが発生すると対話ロボットが誤動作を起こし、カードゲームの継続に重大な支障をきたす。

この問題に対し、本研究ではタスクに無関係な単語を含むガベージクラスをクラス言語モデルに導入し、これにより湧き出し誤りに対処する方法を提案する。

### 3 ガベージクラス

ガベージクラスは、ドメイン外発話がガベージとマッチングすることによってキーワードの湧き出しを防ぐことを目的としている。

ガベージの性能は、登録する単語に依るため、適切な単語選択が重要である。ガベージとして登録する単語を次の 2 つの方法で作成している。

\*Language model construction method for ignoring the speech unrelated to the task. by Mitsuru TAKAOKA, Takuya KAMIHIRA, Hiromitsu NISHIZAKI and Yoshihiro SEKIGUCHI (University of Yamaguchi)

### 3.1 3つ組み音節の利用

母音・鼻音・破裂音・摩擦音の計 35 音節を組み合わせた 3 文字を 1 単語 (例: なかさ) として生成する。生成した単語のうち、キーワードと類似したもの (3 文字中 2 文字以上がキーワードとマッチングするもの) はあらかじめ取り除くことでキーワードと無関係な単語を選択し、ガベージとしている。

### 3.2 日本語話し言葉コーパス (CSJ)[2] の模擬講演から抽出した単語の利用

CSJ の模擬講演からは 70755 単語が抽出できたが、登録数が膨大になるため、本研究では出現頻度を基準とし、頻度 50 回以上の単語を採用し、キーワードを含む単語を削除することで、キーワードと無関係な単語を選び、ガベージとしている。

## 4 評価実験

2つの方法で作成したガベージクラスをクラス言語モデルに導入した評価実験を行った。

### 4.1 実験条件

言語モデル構築時に考慮すべき点は、学習テキスト中の文章構成である。学習テキストには、ドメイン発話の学習としてキーワードのクラス、ドメイン外発話の学習としてガベージクラス、の 2 種類が存在する。テキスト作成時にこれらを混ぜ合わせるが、どちらを多く学習するかに応じて音声認識結果に影響が出る。

このため、複数パターンの学習テキストを用意した。ドメイン発話とドメイン外発話の比率を 10:0 から 10% ずつ変化させた、合計 11 パターンのテキストを作成、それぞれから言語モデルを構築し、音声認識結果を評価する。

言語モデルは、2つの観点 (ドメイン発話を正しく認識できたか / ドメイン外発話を正しく棄却できたか) から評価する。評価用音声は、カードゲームで使用されている発話データをドメイン発話 / ドメイン外発話にそれぞれ分けたもの、各 200 データずつ用意した。

評価はドメイン発話は取得率、ドメイン外発話は棄却率で行う。計算式は以下の通り。

$$\text{取得率 (棄却率)} = \frac{\text{正解発話数}}{\text{全発話数}} \times 100 \quad [\%] \quad (1)$$

式 (1) における正解発話数は、ドメイン発話の評価では、正しくキーワードを認識した発話数であり、ドメイン外発話の評価では、正しく棄却した発話数である。

### 4.2 実験結果

2種類の発話データを使った音声認識結果をグラフ化してまとめたものを図 2, 3 に示す。3つ組み音節をガベージに採用した場合、模擬講演の頻出単語を採用した場合よりもドメイン外発話を棄却しにくく、全体的に棄却率が低い結果となった。このことから、ガベージとして用いる単語がドメイン外発話の棄却に大きく影響していると言える。

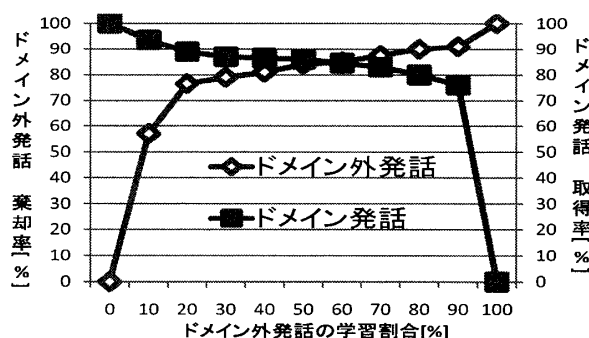


図 2 音節組み合わせをガベージとして利用

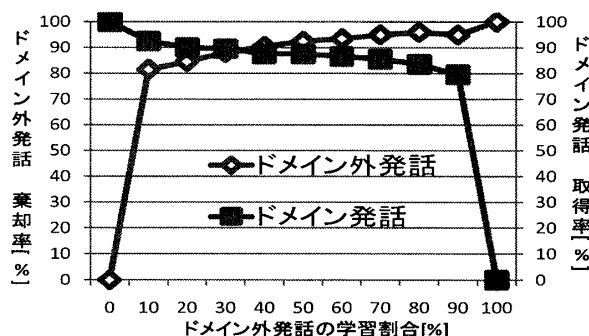


図 3 模擬講演の頻出単語をガベージとして利用

また、図 3 のグラフの交点は、学習テキスト中のドメイン発話とドメイン外発話を 7:3 の比率で学習すると、どちらの発話であっても、正しく音声認識できる可能性が 89% であるということを示している。

以上より、キーワードから構築したクラス言語モデルにガベージクラスを用いることは、ドメイン発話 / ドメイン外発話双方を正しく音声認識させることに非常に有効であることが分かった。

## 5 おわりに

ガベージクラスを導入し、低コストに湧き出し誤りの少ない言語モデルを構築する手法を提案した。

提案手法ではキーワードと無関係な単語を模擬講演から抽出又は 3つ組み音節から生成している。ガベージクラスをキーワードを組み合わせた言語モデルに導入することで、およそ 89% の確率で正しい音声認識ができることを示した。

今後はキーワードと無関係な単語を他のコーパスから得ることや、単語の選別方法を検討することで更なる音声認識能力の向上を図る予定である。

## 参考文献

- [1] 上平拓弥, 手川高彰, 高岡充, 関口芳廣, 西崎博光. 対話ロボットのためのガベージクラスを利用したキーワード認識. 日本音響学会 2009 年秋季研究発表会講演論文集, pp. 185-186, 2009.
- [2] K. MAEKAWA. Corpus of spontaneous Japanese: Its design and evaluation. *Proc. ISCA & IEEE Workshop SSPR 2003*, 2003.