

連続2ターンのユーザ発話を用いた 発話内の未知語有無推定

大塚 嗣巳[†]

中村 友昭[‡]

中野 幹生[‡]

駒谷 和範[†]

[†]名古屋大学大学院 工学研究科 電子情報システム専攻 [‡](株)ホンダリサーチ インスティテュート ジャパン

1. はじめに

音声対話システムにおいて、未知語の誤認識はユーザの意図しない応答生成に繋がる。未知語とは、音声認識用辞書に存在しない単語を指す。未知語の存在を特定することは音声対話システムにおいて重要であり、様々な方法が提案されてきたが、まだ改善の余地がある [1, 2].

人間は一度分からなかった単語を再度聞いても分からない時、それが未知語である可能性を考える。一方、従来の方法では1つの発話だけで未知語かどうかを判定しようとするが、このような推論はしない。我々は人間のような推論を音声対話システムに実装することを目指す。

本研究は2ターンのユーザ発話(発話対)を用いて発話内の未知語有無を推定する。この時、発話対には同じ単語が含まれることを仮定する。認識辞書内の単語として誤認識される未知語を1発話のみから特定することは難しいものの、ユーザ発話を誤認識した場合やユーザに確認する場合に、ユーザが再度同じ未知語を発話することは考えられる(図1上部)。これを利用して推定する。

2. 発話対による発話内の未知語有無の推定

我々は Rastrow らと同様に、未知語有無の推定を二値分類問題として扱う [1]。発話の音声認識結果から抽出した特徴量を用いる(図1下部)。

提案手法では1発話のみから得られる特徴量だけでなく、発話対をペアとして考えたときに得られる特徴量を用いる。未知語は通常、認識辞書内の単語に誤認識されるものの、未知語が含まれる複数の発話の認識結果には共通する傾向があると考えられる。したがって、その傾向を利用できれば、未知語の存在を特定できると考えた。これにより、1発話のみから未知語の有無を推定する従来手法よりも良い推定精度が得られると考えられる。

特徴量は、具体的には発話対の各発話から得られる特徴量の差や平均を用いる。これは、認識辞書内単語と未知語それぞれが使用された発話対の認識結果に生じると考えられる差分を抽出するためである。例えば、辞書内単語は発話対で同じ単語が認識されるが、未知語は異なる単語に認識される可能性がある。また、辞書内単語の単語信頼度は発話対で共に高く差が小さいが、未知語の場合、少なくとも一方の単語信頼度が低く出力され、差が大きくなる可能性が考えられる。

3. 実験

3.1 実験データ

実験データとして、被験者がシステムに物体や人の名前を教える対話を収録したものをを用いた。被験者は男性7名、女性11名の計18名である。データは合計286発話対で、被験者がシステムに教えた単語(名前と呼ぶ。図1ではペンチ)の異なり数は223単語である。発話対

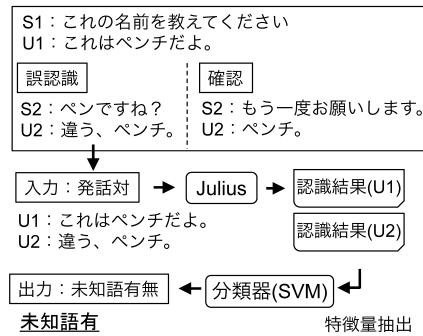


図1: 対話例およびシステム概要 (S: システム, U: ユーザ)

表1: データ分割方法

	data1	data2	total
言語モデル作成用データ	71	71	142
未知語用データ	68	76	144
total	139	147	286

と名前の数が異なるのは、異なる被験者が同一の名前を発話するケースが存在したためである。

対話データを表1のように4分割した後、言語モデル作成用データと未知語用データの両方を含む発話データ集合を data1, data2 の2種類作成した。なお、4種類のデータ集合は、被験者、名前が重複しない用に分割した。

3.2 特徴量

Julius^{‡§}の音声認識結果から76種類の特徴量を抽出した。基本となる表2の15種類の特徴量を表3で表される a~e の5種類の特徴量生成パターンで求めた。また、発話対の各発話から認識された名前の音素表記の編集距離も利用した。生成パターンのうち、d, e が発話対を考慮した特徴量である。

特徴量抽出のために2種類の言語モデルを使用した。言語モデル作成用データから作成された小語彙のクラス 3-gram のものと、Julius 付属の dictation-kit 内で使用されている大語彙 3-gram のものである。小語彙のモデルには名前に対し [物体名] というクラスを付与する。特徴量 M, N は大語彙の言語モデルを、特徴量 O は両方を、それ以外は小語彙のものを使用した。

発話検証スコア(O)とは、発話全体の認識結果の信頼度に相当するスコアである。対象とする言語モデルと大語彙言語モデルそれぞれを用いて得られた音声認識結果の音響スコアの差を話速で割って算出する。

編集距離は、発話対の各発話内で出現した名前の音素表記から算出する。発話対内の未知語は各発話で異なる単語に認識される可能性があるという直感に基づきこれを用いる。発話対のうち、一方のみにしか名前が存在しない場合、その名前と空文字列との編集距離を計算する。

Estimating Existence of Unknown Word in User Utterances by Considering Repeated User Utterance Pairs: Tsugumi Otsuka (Nagoya Univ.), Tomoaki Nakamura, Mikio Nakano (Honda Research Institute Japan Co., Ltd.) Kazunori Komatani (Nagoya Univ.)

[‡] <http://julius.sourceforge.jp/>
[§] Ver. 4.2.1 を使用した

表 2: 1 発話から得られる特徴量

	特徴量
A	認識単語数
B	話速 (A/発話長)
C	認識スコア
D	言語スコア
E	音響スコア
F	単語信頼度の最大値
G	単語信頼度の最小値
H	単語信頼度の平均値
I	[物体名] クラスの単語信頼度
J	音素数
K	D/B
L	E/B
M	大語彙言語モデルでの言語スコア
N	大語彙言語モデルでの音響スコア
O	発話検証スコア ((E-N)/B)

表 3: 特徴量生成パターン

パターン	生成方法
a	1 発話目の 1 位の認識結果から得られるスコア
b	1 発話目の 1 位と 2 位の認識結果から得られるスコアの差
c	2 発話目の 1 位の認識結果から得られるスコア
d	a と c との差
e	a と c との平均

3.3 実験条件

以下の 4 種類の実験方法を比較した。

- 提案手法 (76 種) :
特徴量 A~O × a~e の 5 パターン + 編集距離
- 1 のうち発話対を考慮した特徴のみ使用 (31 種) :
特徴量 A~O × d, e の 2 パターン + 編集距離
- 1 発話目のみ使用 (30 種) :
特徴量 A~O × a, b の 2 パターン
- 発話対の各々を独立した 1 発話として使用 (30 種) :
特徴量 A~O × a, b の 2 パターン

1, 2 が提案手法, 3, 4 がベースライン手法である。1 は発話対のそれぞれの認識結果から得られた特徴量も使用するのに対し, 2 は発話対を考慮した特徴量のみ使用する。2 で評価を行う理由は, ベースライン手法と使用する特徴量の数に差が無い場合での比較を行うためである。4 は, 発話対の各発話をそれぞれ独立した 1 発話と見なすものである。つまり, 3 と比べて発話数を 2 倍にした場合に相当する。4 を用いる理由は, 提案手法とベースライン手法で同じデータ量を使用した比較を行うためである。ベースライン手法では 30 種類の特徴量を使用する。これは, A~O の 15 種類の特徴量を, 1 発話から生成できる発話生成パターン a, b により算出する。分類器の学習とテストは Weka¹⁾ の SMO (SVM の一実装) をデフォルトのパラメータを用いて使用した。

評価方法には未知語有無の二値分類における正解率を用いた。data1, data2 の一方を学習データ, 他方を評価データとし, 2 回分の正解率の重み付き平均で比較する。

¹⁾ <http://www.cs.waikato.ac.nz/ml/weka/>

²⁾ Ver. 3.6.10 を使用した

表 4: 各手法による未知語有無の推定正解率

実験方法	学習 : data1 評価 : data2	学習 : data2 評価 : data1	重み付き平均
1	77.3%	83.6%	80.4%
2	75.8%	82.0%	78.8%
3	67.4%	71.4%	69.4%
4	73.1%	73.8%	73.4%

表 5: 特徴量を除いた時の正解率の変化

有効な特徴量		有効でない特徴量	
特徴量	正解率増減	特徴量	正解率増減
a-O	-1.99	a-A	1.20
a-J	-0.40	a-D	1.20
a-M	-0.40	a-K	1.20
b-C	-0.40	b-F	1.20
c-A	-0.40	b-I	1.20
d-O	-0.40	b-L	1.20
b-J	-0.38	e-A	1.20
c-O	-0.38	e-D	1.20
編集距離	-0.38	d-K	0.80
e-O	-0.36	d-J	0.78

3.4 実験結果

評価の結果, 発話対を考慮することの有効性を確認した。各実験方法での発話内の未知語有無の推定正解率を表 4 に示す。まず, ベースラインである 3 と 4 を比較した時, 4 の方が推定の正解率が高い。これは学習データの量の差が影響していると思われる。次に, 提案手法である 1, 2 とベースライン手法 4 を比較した時, 1, 2 の方が推定正解率が高い。この結果から, 単に 1 発話のみを使用するだけでなく, 発話対を考慮することが有効であることが確認できた。また, 発話対を考慮した特徴量が, 1 発話のみから得られる特徴量に比べて有効であったことも確認できた。

3.5 有効な特徴量

未知語有無の推定にどの特徴量が影響しているかを調査した結果を表 5 に示す。表 5 は提案手法の 76 種類の特徴量のうち, 推定正解率の増減に最も大きく影響した特徴量を 10 種類ずつ示したものである。正解率増減の列は, 特徴を削除した後の正解率から全ての特徴量を使用した時の正解率 80.6% を引いた値である。正解率増減は, 値が小さいほど有効な特徴量であることを意味する。

発話検証スコア (O) や編集距離は有効であるという結果になった。認識辞書内の名前が発話された場合, 発話対で同じ名前が認識され, 編集距離が小さくなる傾向があり, 未知語の場合は発話対で異なる単語に認識され, 編集距離が大きくなる傾向があるためだと考えられる。

4. まとめ

本研究はユーザ発話内の未知語有無を推定することを目指しており, 本稿では発話対を用いることにより, 1 発話のみの時に比べて推定精度向上が可能であることを報告した。今後は, 同じ名前が繰り返される仮定をしなくても良い方法を検討する。

参考文献

- [1] Ariya Rastrow, Abhinav Sethy, and Bhuvana Ramabhadran. A new method for OOV detection using hybrid word/fragment system. *ICASSP*, pp. 3953–3956, 2009.
- [2] Burget, Lukas, et al. Combination of strongly and weakly constrained recognizers for reliable detection of OOVs. *ICASSP*, pp. 4081–4084, 2008.