

未知語の存在を考慮したユーザ発話のドメイン推定

高橋 裕己[†] 中野 幹生^{‡§} 菅野 重樹[†]

[†] 早稲田大学 創造理工学部

[‡] 早稲田大学 理工学術院

[§] (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

航空券の予約のような単一のドメインの対話を扱う対話システムに対し、複数のドメインの対話を行う対話システムはマルチドメイン対話システムと呼ばれ、近年盛んに研究されている。特に、オフィスや家庭などで人間と対話しながら様々な種類のタスクを行う対話ロボットは、マルチドメイン対話機能が必須である。例えば、物を運んだり、物を取ってきたり、スケジュールなどの情報検索を行えるような対話ロボットは、物の運搬要求を理解する対話ドメイン、物を取ってくる要求を理解する対話ドメイン、情報検索要求を理解する対話ドメインを扱える必要がある。

マルチドメイン対話システムにおいては、ユーザの各発話がどのドメインの発話かを判定する必要がある。ドメインを推定することで、ドメインに応じた確認や聞き返しを行うことができる。しかしながら、ユーザ発話中に未知語、すなわち音声認識・理解部に登録されていない語があると、音声認識・理解の誤りが起こり、ドメインの判定が難しくなる。特に、ロボットに物体の名前を教えるような対話ドメインでは、未知語が発声される確率が高い。本稿では、音声認識信頼度や文型との適合性などの特徴量を用いて未知語検出とドメイン選択を同時に行う方法を提案する。

2 ドメイン選択

一般的にドメイン選択には二つの方法がある。一つは分類問題として扱う方法である。これは、 n 個の対話ドメインがあったとき、入力発話がどのドメインなのかを n 分類問題として解く方法である [2]。分類器は、正解ドメインがアノテートされている訓練データを用いて学習する。どのドメインにも属さない発話を検出するときには、「その他」ドメインを加え、 $n+1$ 分類問題にする。分類器で用いる特徴量には、単語の出現頻度やドメイン毎の文型の一致度を用いる。

もう一つは、入力発話がどのくらい各ドメインの発話である可能性が高いかを表すスコアをドメイン毎に算出し、最も高いスコアを出したものを選ぶ方法である [3]。スコアの算出はドメイン毎に異なる関数によって行う。この関数も訓練データによってパラメータ調整を行う。

本研究では後者の方法を採用する。後者の方法は、ドメイン毎に異なる特徴量を用いることができることや、ドメインが増えたときに訓練データセットを用意しな

おさなくても、新しいドメインの訓練データのみを用意すればよいことなどの利点がある。

3 未知語を含む発話のドメイン選択

従来のドメイン選択手法は、未知語、すなわち、システムの音声認識・言語理解語彙に登録されていない単語が現れることを考慮していない。未知語が現れると音声認識・言語理解に失敗する。そのため、ドメイン選択でよく用いられる、認識結果に表れる単語集合や、発話検証スコア（ドメイン依存の言語モデルを用いた場合の音響スコアと大語彙統計言語モデルを用いた場合の音響スコアの差を発話長で割ったもの）[1] などが有効とは限らない。これは、ロボットに物の名前を教えるようなドメインの場合、顕著に問題が生じる。

そこで、未知語を陽に扱う特徴量を用いることで、ドメイン選択の精度を上げる方法を提案する。具体的には、未知語の有無にかかわらず、音声認識結果が特定の文型にマッチするかどうかを判定し、その結果を特徴量として用いる。例えば、「これは〈A〉です」のような文型を用いる。〈A〉の部分にはどのような単語列がマッチしても良いとする。これにより、そのドメインに表れやすい文型かどうかを判断することができる。また、文型にマッチした場合、〈A〉の部分に現れる単語の単語信頼度を見ることで、未登録語を含むかどうかを推定することができる。

4 実験

4.1 概要

提案手法の有効性を調べるため実験を行った。本実験のタスクはロボットに物体の名前を教えるドメインの発話を、他のドメインの発話と区別することである。物体の名前を教えるドメインの発話では、その名前が未知語の場合もあると仮定する。本実験では、入力を未知語を含むドメイン内発話、未知語を含まないドメイン内発話、ドメイン外発話の3クラスに分類する実験を行った。

4.2 データ

本実験では、13人分の発話データ、1,197発話を用いた。物の名前を教える発話（本実験ではこれをドメイン内発話とする）が792発話、その他のドメインの発話（ドメイン外発話、物の名前を尋ねる発話、天気情報を聞く発話、電話番号を聞く発話）が405発話である。ドメイン内発話はすべて定型的な発話である。

Utterance Domain Selection that Takes the Existence of Out-of-Vocabulary Words into Account: Yuki Takahashi (Waseda Univ.), Mikio Nakano (Waseda Univ./Honda Research Institute Japan Co., Ltd.), and Shigeki Sugano (Waseda Univ.)

表 1: 実験で用いた特徴量

F1	発話長
F2	音素数
F3	発話速度 (F2/F1)
F4	信頼度の平均値
F5	信頼度の最小値
F6	文型にマッチしたとき、名前部分にマッチした単語の信頼度の平均
F7	音声認識結果の単語数
F8	音響スコア
F9	言語スコア
F10	F8/F1
F11	F9/F1
F12	大語彙言語モデルを用いた音声認識結果の音響スコア
F13	発話検証スコア ((F8-F12)/F1)
F14	文型にマッチしたかどうか (0 か 1 のバイナリ値)
F15	文型とマッチした時のスコア (名前部分にマッチした単語列が長いほどスコアが低くなる)

(F12, F13 以外は、ドメイン用言語モデルを用いた音声認識結果から抽出.)

4.3 音声認識

音声認識には、Julius ver. 4.2.1*を用いた。音響モデルは、Julius ディクテーション実行キット 4.2 付属のものを用いた。言語モデルは、ドメイン用言語モデルと、大語彙言語モデルの両方を用いた。ドメイン内発話のうち 361 発話を用い、palmkit 1.0.32 を用いて学習した。この言語モデルに含まれる単語は既知語となり、それ以外は未知語となる。このプロセスで、ドメイン内発話のうち既知語を含むものと未知語を含むものが同数になるように、言語モデルの学習時に、いくつかの発話の物体の名前の部分を人為的に変更した。大語彙言語モデルは Julius ディクテーション実行キット 4.2 付属のものを用いた。

4.4 特徴量

表 1 に示した特徴量を用いた。ドメイン用言語モデルを用いた音声認識結果、大語彙言語モデルを用いた音声認識結果とも、最上位 (1-Best) の音声認識結果を用いて特徴量を計算した。信頼度は Julius の単語信頼度を用いた。

4.5 分類結果

分類器は Weka ver.3.6.8[4] の Logistic(ロジスティック回帰) を用いた。パラメータはデフォルト値を用いた。10 分割交差検定による分類器の評価結果を表 2 に示す。特徴量選択は Weka の ClassifierSubsetEval を用いて行った。

4.6 考察

未知語処理に関する特徴量を用いた場合の方が分類精度が高いことから、提案手法は有望であると考えられる。しかしながら、その差はわずかである。これは、名前を教える発話が定型的なものであったため、言語モデルへの適合度が高く、たとえ未知語があったとしても、ドメイン選択が容易であったためだと考えられ

表 2: 分類結果

条件	正解 \ 推定	OOD	IDOOV	IDIV	分類精度
A	OOD	359	43	3	83.96%
	IDOOV	41	338	52	
	IDIV	2	51	308	
B	OOD	359	44	2	83.96%
	IDOOV	42	337	52	
	IDIV	2	50	309	
C	OOD	365	38	2	85.38%
	IDOOV	30	348	53	
	IDIV	3	49	309	
D	OOD	370	33	2	85.96%
	IDOOV	33	348	50	
	IDIV	5	45	311	

条件 A: 未知語処理 (特徴量 F6, F14, F15) なし, 特徴量選択なし

条件 B: 未知語処理なし, 特徴量選択あり
特徴量選択により削除された特徴量: F11

条件 C: 未知語処理あり, 特徴量選択なし

条件 D: 未知語処理あり, 特徴量選択あり
特徴量選択により削除された特徴量: F1, F2, F4, F10, F11, F14

OOD: ドメイン外

IDOOV: ドメイン内, 未知語あり

IDIV: ドメイン内, 未知語なし

る。文型のバリエーションが高い場合の調査が必要である。

5 おわりに

本稿では、未知語が高い頻度でユーザ発話に現れるようなドメインにおいてドメイン推定の精度を高める方法を提案した。

本稿で記述した実験は予備的なものであり、今後、より現実的な設定で実験を行う。具体的には、ドメイン内発話として、自由な文型で発話したものをを用いる。また、言語モデルのトレーニングデータ、マッチさせる文型を構築する際に参考にするデータをテストデータと別にする。また、ドメイン内外判定だけでなく、ドメイン内発話であるスコアも算出し、マルチドメイン対話システムに組み込んだ場合に本手法が有効であるかどうかの確認を行う。

謝辞 本研究の初期段階において、Brian Romanowski 氏、成松宏美氏に御協力を頂きました。深謝いたします。

参考文献

- [1] K. Komatani et al.: Introducing utterance verification in spoken dialogue system to improve dynamic help generation for novice users. In *Proc. SIGdial*, 2007.
- [2] I. R. Lane et al.: Topic classification and verification modeling for out-of-domain utterance detection. In *Proc. Interspeech-2004*, 2004.
- [3] M. Nakano et al.: A two-stage domain selection framework for extensible multi-domain spoken dialogue systems. In *Proc. SIGDIAL*, 2011.
- [4] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann, 2005.

*<http://julius.sourceforge.jp/>