

5M-1

# 複数ドメイン音声対話システムにおける 対話履歴を利用したドメイン選択の高精度化

神田 直之<sup>†</sup> 駒谷 和範<sup>†</sup> 中野 幹生<sup>‡</sup> 中臺 一博<sup>‡</sup> 辻野 広司<sup>‡</sup> 尾形 哲也<sup>†</sup> 奥乃 博<sup>†</sup>

<sup>†</sup> 京都大学大学院 情報学研究科 知能情報学専攻      <sup>‡</sup> (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

我々は、シングルドメインの音声対話コンポーネントを統合することにより、複数のドメインを扱うことのできるシステム(マルチドメイン音声対話システム)を開発した。マルチドメイン音声対話システムでは、ユーザ要求がどのドメインでなされているのかを推定する処理(ドメイン選択)が必要である。音声対話システムでは音声認識誤りが不可避であるため、誤りを含んだ音声認識結果からも正確にドメインを推定できる頑健さが要求される。これまで、ユーザ発話音声認識結果から応答すべきドメインを推定する研究が行われてきた [1, 2]。これらの研究では、ひとつの発話から得られる情報のみを用いて応答すべきドメインを推定している。しかしながら、ドメインの選択はシステムとユーザのやりとりの中で決定されるものであり、このような手法は十分ではない。本研究では対話の状態や進行から得られる情報も利用することで、より音声認識誤りに頑健なドメイン選択を行う。

## 2. マルチドメイン音声対話システムのアーキテクチャ

マルチドメイン音声対話システムでは単一ドメインのシステムよりも複雑なシステム設計が必要となるため、各ドメインが独立に設計できる分散型 [3] のアーキテクチャが望ましい。ドメインごとに独立して記述できる部分とドメイン間の関連を考慮する必要のある部分を切り分ける。後者の部分を最小化することにより、システム設計者は個々のドメインを半独立的に設計することができ、各ドメインの改良や追加が容易となる。

本研究でもこの流れに立ち、我々が開発した分散型のシステムアーキテクチャ [4] に基づきシステム設計を行った。システムは大きくわけて、各ドメインでの対話を担当するエキスパートと、それを統合するシステム本体に分かれる(図1)。システム本体は主に、どのドメインエキスパートにユーザ発話の処理を行わせるかの決定を行い、言語理解や対話状態の変更、システムの行う対話行為の決定といった処理は各ドメインのエキスパートが

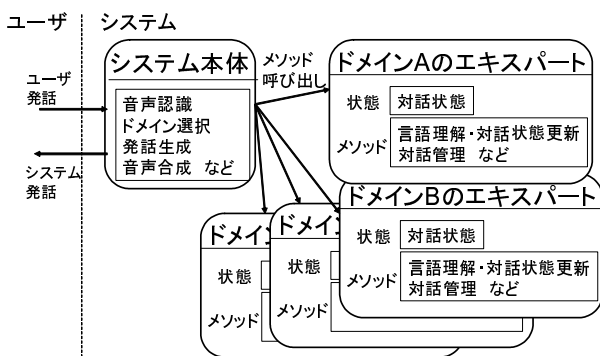


図1: マルチドメイン音声対話システムの構成

Robust Domain Selection using Dialogue History in Multi-Domain Spoken Dialogue Systems: Naoyuki Kanda, Kazunori Komatani (Kyoto Univ.), Mikio Nakano, Kazuhiro Nakadai, Hiroshi Tsujino (HRI-JP), Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

U1: 法輪寺の住所(寺社案内ドメイン)  
S1: 法輪寺の住所は~です。  
U2: 法輪寺の料金(寺社案内ドメイン)  
(音声認識誤り「ホリデーインの料金(ホテルドメイン)」と誤認識)  
S2: ホリデーインの料金は~です。  
U3: 法輪寺の料金(寺社案内ドメイン)  
(再び「ホリデーインの料金」と誤認識)  
  
S3 誤: ホリデーインの料金は~です。  
S3 正: ホテルの情報についてお尋ねですか?  
  
U4: いいえ  
S4: 寺社の案内に戻ってよろしいですか。

図2: ドメイン選択(III)の場合の対話例

担当する。

このように分散型のアーキテクチャでは、ユーザ発話の処理は各ドメインでのエキスパートに任せてシステム本体は関知しないため、どのドメインにユーザ発話の処理を任せるかの判断(すなわちドメイン選択)が重要となる。また、分散型アーキテクチャの特徴である、新たなドメインの追加・変更の容易さを確保するためには、ドメイン選択で利用する情報はドメインに依存せず利用できるものでなければならない。

## 3. 対話履歴を利用したドメイン選択

### 3.1 ドメイン選択の課題

本研究では、対話の文脈や進行から得られる情報を利用して、ドメイン選択を行う。同様の研究に、ひとつ前の応答を行ったドメインを維持する制約を設けてドメインの選択を行う研究がある [3, 5, 6]。これらの手法では、ひとつ前の応答を行ったドメインが正しく推定されていた場合、音声認識誤りによる誤ったドメイン遷移を防ぐことができる。しかしながら、ひとつ前の応答を行ったドメインが誤っていた場合、その誤ったドメインを連続して選択してしまう問題が生じる。この問題は、ひとつ前の応答を行ったドメインを無条件に信頼しているために生じる。本研究では、ひとつ前の応答を行ったドメインの履歴や状態も考慮してドメイン選択を行うことにより、この問題の解決を図る。

また、ドメイン選択誤りが連続する場合には、ひとつ前の応答を行ったドメインと音声認識結果の双方が誤りである、という状況が生じる(図2のU3)。この検出は、ドメインは正しいが音声認識が誤っているのか、ドメインそのものが誤っているのかを判断しなければならないため、単純に音声認識誤りの連続を検出するよりも難しい問題であり、これまで扱われていない。そこで本研究ではドメイン選択問題を、応答すべきドメインが、(I) ひとつ前の応答を行ったドメイン、(II) 音声認識結果に対する最尤のドメイン、(III) その他のいずれかのドメイン、のいずれであるかを選択する問題と捉える(図3)。選択肢(III)が検出できれば、誤りと判断されたドメインへの遷移を防いだり、以前のドメインに戻るなどの回復戦

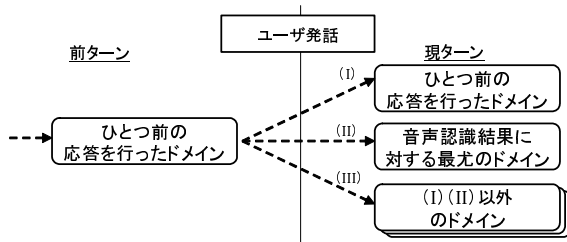


図 3: ドメイン選択の概略

略を採ることができる(図2のU3正,及びU4)。

### 3.2 ドメイン選択器の構成

本研究では応答すべきドメインが,前節の選択肢(図3)のいずれであるかを判別する判別器を,対話データから学習する.対話データには発話ごとに(I)~(III)のいずれかを人手によってラベル付けする.これを下記の特徴量を用いて判別する.

本研究では図3に対応してひとつ前の応答を行ったドメインの履歴・状態を表現する特徴量と,ユーザ発話音声認識結果に関する特徴量,及び各ドメイン選択を行った場合にどのような履歴・状態になるかを表現する特徴量を利用した[7].ひとつ前の応答を行ったドメインの履歴・状態を表現する特徴量を用いることで,そのドメインが誤っている可能性も包含してドメイン選択を行う.ここではユーザから肯定・否定の応答があった回数や,スロットが変化した割合などを利用した.ユーザ発話音声認識結果に関する特徴量は,従来から用いられている特徴量に相当する.ここでは各ドメインで言語理解した音声認識結果の音声認識スコアや,単語信頼度の平均などを利用した.各ドメイン選択を行った場合にどのような履歴・状態になるかを表現する特徴量は,音声認識結果を各ドメインエキスパートが言語理解した後にシステム本体に返す情報である.各ドメイン選択後のタスクの状態や,そのドメイン選択をした時のスロット変化などを特徴量として利用した.

## 4. 評価実験

### 4.1 評価用データの収集

提案手法を評価するために,5ドメインのマルチドメイン音声対話システムを実装した.作成したドメインは,レストランデータベース検索,ホテルデータベース検索,寺社案内,天気案内,バス運行情報案内である.音声認識エンジンはJulian[8]を用いた.音声認識文法の語彙サイズは7,373である.

上記のシステムを用いて,10名の被験者から対話データを収集した.被験者はまず,音声入力のタイミングに慣れるため簡単なシナリオに基づき10分ほど練習を行った.その後ドメインを3~4回変更することを想定した状況シナリオに基づいて対話を行った.同様の条件で3対話を行った.今回のシステムでは,人手で作成したルールによってドメイン選択を行った.

実験により得られた発話は総計2205発話(221発話/人,74発話/対話)また,単語正解率は63.3%であった.

### 4.2 ドメイン選択の精度の評価

以下の2種類の方法で,ドメイン選択の精度を比較評価した.

従来手法: 音声認識結果を受理できたドメインのうち,音響尤度が最大のものを選択する.ただし,ひとつ前の応答を行ったドメインには音響尤度に値 $\alpha$ を加算して比較する. $\alpha$ は複数の値を評価検討した.

表 1: 各手法におけるドメイン選択の誤り数/発話数

正解	ひとつ前のドメイン	最尤のドメイン	その他のドメイン	計
従来手法	162/1451	247/526	228/228	637/2205
提案手法	130/1451	288/526	145/228	563/2205

提案手法: 提案するドメイン選択.今回,ドメイン選択器をC5.0[9]により構成した.

評価はいずれも発話ごとに行った.また,同一スコアのドメインが複数存在した場合,その中からランダムにひとつのドメインを選択して正解判定を行った.

従来手法では $\alpha$ の値が大きいくほど,ひとつ前の応答を行ったドメインを維持する制約が大きくなる. $\alpha=45$ としたときに,誤り数が最も少なくなり,637であった.このときのドメイン選択誤り率は28.9%(=637/2205)であった.また,正解ラベルが「(I)ひとつ前の応答を行ったドメイン」「(II)音声認識に対する最尤のドメイン」のどちらでもないものが228個含まれていた.これらは従来手法では正解を選択することができない発話である.

次に,提案手法による評価を行った.ここでは,5-fold cross validationにより評価結果を得た.C5.0のカットオフパラメータを最適化した結果,カットオフパラメータが25の時にドメイン選択誤り数が最小の563となった.この時のドメイン選択誤り率は25.5%(=563/2205)であった.従来手法からのドメイン選択誤り削減率は11.6%(=74/637)である.これらを正解ラベルごとに分類したものを表1に示す.改善が大きかったのは,正解ラベルが「(III)その他のいずれかのドメイン」であるものであり,誤り数が従来手法に比べ83減少し,145となった.これは,音声認識結果もひとつ前に応答したドメインも誤りであるという状況の約1/3において,回復戦略を採るのがよいことを検出できたことを表している.

## 5. まとめ

本研究ではマルチドメイン音声対話システムにおいて,対話の状態や履歴を用いることでより音声認識誤りに頑健にドメイン選択する手法について述べた.被験者10名による評価実験では,本研究で提案した手法により従来手法に比べドメイン選択誤りが11.6%削減されることを確認した.

謝辞 本研究の一部は,学振科研費,21世紀COEプログラムの支援を受けた.

## 参考文献

- [1] T. Isobe et al. A study on domain recognition of spoken dialogue systems. In *Proc. EUROSPEECH*, 2003.
- [2] 安田宜仁,他.単一ドメインシステムの統合による複数ドメイン音声対話システム.情処研報,2003-SLP-45-20,2003.
- [3] B. Lin et al. A distributed agent architecture for intelligent multi-domain spoken dialogue systems. In *IEICE Trans. on Information and Systems*, E84-D(9), pp. 1217-1230, 2001.
- [4] M. Nakano et al. A two-layer model for behavior and dialogue planning in conversational service robots. In *Proceedings of IROS-2005*, pp. 1542-1548, 2005.
- [5] 長森誠,他.マルチドメイン音声対話システムの構築手法.情処研報,2000-SLP-31-7,2000.
- [6] I. O'Neill et al. Cross domain dialogue modelling: An object-based approach. In *Proc. ICSLP*, Vol. 1, 2004.
- [7] 神田直之,他.複数ドメイン音声対話システムにおける対話履歴を利用したドメイン選択の高精度化.情処研報,2006-SLP60-11,2006.
- [8] 河原達也,李晃伸.連続音声認識ソフトウェアJulius.人工知能学会誌,Vol. 20, No. 1, pp. 41-49, 2005.
- [9] C5.0. <http://rulequest.com/index.html>.