

## 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

naoyuki@kuis.kyoto-u.ac.jp {komatani,ogata,okuno}@i.kyoto-u.ac.jp

<sup>‡</sup>(株) ホンダ・リサーチ・インスティテュート・ジャパン

{nakano,nakadai,tsujino}@jp.honda-ri.com

複数のドメインを扱う音声対話システムにおいて、対話の文脈や進行に関する特徴量を導入してより精度よくドメイン選択を行う手法を開発したので報告する。本稿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) それ以外のいずれかのドメイン、のどれに該当するかを判別する問題と捉える。ドメイン選択の正解を与えた対話データから、対話の文脈や進行に関する特徴量を用いて上記を判別する決定木を学習することにより、ドメイン選択器を構成した。5 ドメインのマルチドメイン音声対話システムを用いた 10 名の被験者による評価実験の結果、音声認識尤度に基づく従来のドメイン選択手法に比べ、ドメイン選択誤りが 11.6% 削減された。

## Robust Domain Selection using Dialogue History in Multi-Domain Spoken Dialogue System

NAOYUKI KANDA<sup>†</sup>, KAZUNORI KOMATANI<sup>†</sup>, MIKIO NAKANO<sup>‡</sup>, KAZUHIRO NAKADAI<sup>‡</sup>,  
HIROSHI TSUJINO<sup>‡</sup>, TETSUYA OGATA<sup>†</sup> and HIROSHI G. OKUNO<sup>†</sup>

<sup>†</sup> Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

<sup>‡</sup> Honda Research Institute Japan Co., Ltd.

We have developed a robust domain selection method using dialogue history in multi-domain spoken dialogue systems. We define domain selection as classifying problem among (I) the domain in the previous turn, (II) the domain in which N-best speech recognition results can be accepted with the highest recognition score, (III) other domains. We constructed a classifier by decision tree learning with dialogue corpus. The experimental result using 10 subjects shows that our method could reduced 11.6% domain selection error, compared with a conventional method using speech recognition likelihoods only.

### 1. はじめに

これまで、バス運行情報案内やレストラン検索など様々なタスクドメインにおいて音声対話システムが作成されてきた<sup>1),2)</sup>。これらのシステムの多くは扱えるドメインが、バス運行情報のみ、レストランのみといったように 1 種類に限られており、ユーザはそれらを意図的に使いわける必要がある。これに対し本稿では、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）の開発を行う。ひとつのシステムで複数のドメインを扱うことにより、ユーザの要求が複数のドメインをまたがるような、複雑なタスクも扱うことができる。またドメイ

ン間で履歴を共有することにより、シングルドメイン個々を利用するよりもスムーズなタスク達成が可能となる。

マルチドメイン音声対話システムでは、まず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。音声を扱う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。最も単純なドメイン選択方法は「バス」「レストラン」といったようにユーザに明示的にドメイン名を発話させることである。しかしながらこの方法は、対話が冗長で自然さに欠けたものとなる上に、システム設計者の定義したドメインの区切りをユーザが理解することを要求され、ユーザに強い負担を強いる

ことになる．このため、自然なユーザ発話から応答すべきドメインを推定することが行われてきた<sup>3)~5)</sup>．これらの研究では、ひとつの発話音声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、このような推定手法は十分ではない．本稿では対話の文脈や進行に関する特徴量も考慮することで、より音声認識誤りに頑健なドメイン選択を行う．

本稿では、マルチドメイン音声対話システムにおいて、対話の文脈や進行に関する特徴量(対話的特徴量)を導入してより精度よくドメイン選択を行う手法を提案する．本稿ではドメイン選択問題を、対話的特徴量とユーザ発話から、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)それ以外のいずれかのドメイン、のどれに該当するかを判別する問題と捉える．正解を付与した対話データから、上記を判別する決定木を学習することにより、ドメイン選択器を構成する．

## 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．各ドメインが密接に関連付けられた設計がなされていた場合、あるドメインの改変の影響がシステム全体に及ぶため、ドメインを作成し直したり、新たなドメインを追加することが困難となる．そこで、各ドメインが独立に設計できる分散型<sup>6)</sup>のアーキテクチャが提案されている．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<sup>4),6)~11)</sup>．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．

本稿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<sup>12)</sup>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)．ユーザ発話が入力されると、まずシステム本体で音声認識を行い、結果を各エキスパートに送信する．各ドメインを担当するエキスパートでは音声認識結果から言語理解、対話状態の更新を行い、ドメイン選択に必要な情報をシステム本体に返す．システム本体では各エキスパートから送信された情報を基に、応答すべきドメインの選択を行う．選択されたドメインを担当するエキスパートは、対話状態に基づき次に行う対話行為を決定する．システム本体は対話行為から発話生成、音声合成を行い、発話合成

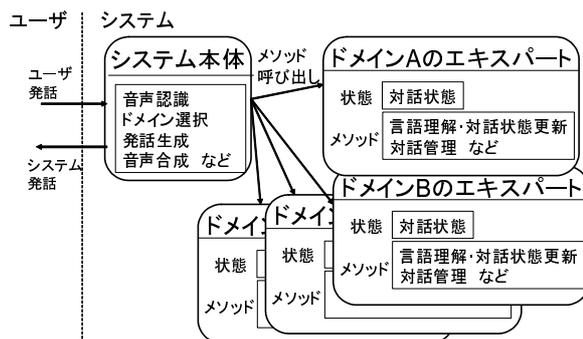


図1 マルチドメイン音声対話システムの構成

音声としてユーザに出力する．システム本体とエキスパートとの通信は、システム本体から各エキスパートのメソッドを呼び出す形で行われる．ドメインエキスパートの機能要件として、エキスパートは発話理解や応答選択などのメソッド<sup>12)</sup>を持つことが要求される．

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がドメイン非依存のものである必要がある．

## 3. ドメイン選択の課題

本稿では、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．同様にドメイン選択の際に対話から得られる情報を用いたものとして、ひとつ前の応答を行ったドメインの情報を利用した研究がある．文献6),7)では前ターンのドメインにより近いドメインを選択するような制約を設けている．さらに文献9)では、タスク達成するまではドメイン遷移を行わない．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、音声認識誤りや言語理解誤りによって起こ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、ひとつ前で推定され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、ひとつ前で推定されたドメインを無条件に信頼しているために生じる．本稿では、前ターンで推定され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、音声認識結果が正しくない時に正しいドメインを維持するためには、音声認識結果とドメイン選択問題を分離する必要がある．これまでに述べた従来手法では(ひとつ前のドメインを利用するものも含めて)いずれも、応答すべきドメインの候補が、その音声認識結果を解釈できるエキスパート(もしくは同等のもの)を持つドメインに限定されていた．しかしながら

ドメイン選択の際に選択されなかったドメインの対話状態は対話状態変更以前に戻す処理も行っている．

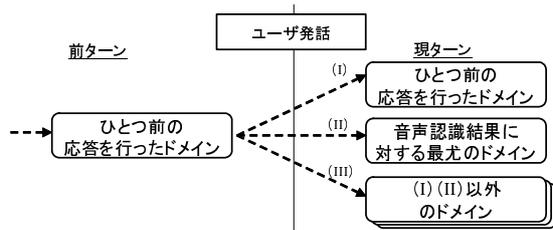


図2 ドメイン選択の概略

U1: 河原町の居酒屋 (レストランドメイン)  
 S1: 河原町の近くの居酒屋を検索します。30件見つかりました。  
 U2: 玉の光が飲めるところ (レストランドメイン)  
 (玉の光は未知語「丹波橋の名所を(観光ドメイン)」と誤認識)  
 S2 誤: 丹波橋の名所を検索しました。10件あります。(観光ドメイン)  
 S2 正: 発話が理解できませんでした。その他の条件はありませんか?(レストランドメイン)

図3 ドメイン選択 (I) の場合の対話例

音声認識誤りが生じた結果、正しいユーザ要求ドメインのエキスパートでその音声認識結果を解釈できなくなる、という事態が(特に未知語発話、文法外発話の場合に)存在する。そこで本稿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) それ以外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2)。

音声認識結果とドメイン選択問題を分離することで、音声認識が誤った場合でも、ひとつ前に応答したドメインが正しければ、それを維持することができる。この対話例を図3に示す。ここではまずU1でユーザがレストランに関する発話を行う。次にU2でユーザがやはりレストランに関する発話を行ったが、未知語を含む発話であったため、レストランドメインで理解できる音声認識結果が得られなかった。この場合、U2に対する応答はレストランドメインで行うべきであるが、従来手法ではU2の音声認識結果を受理できた寺社ドメインへ遷移してしまう(S2誤)。このような場合でも、ドメイン選択(I)が選択されることで「S2正」のようにレストランドメインで応答を続けることができる。

また(III)を定義し、そのような場合を同定することで、前ターンや音声認識結果に基づくドメインエキスパートに制御権を与えるという通常の制御を行ってはいけな場合を検出できる。これは誤りが続く場合に相当するが、ドメインは正しいが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点で、単純に誤りの連続を検出するよりも難しい問題である。この例を図4に示す。U1で寺社案内ドメインに関する発話を行っていたユーザが

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4 ドメイン選択 (III) の場合の対話例

再びU2で寺社案内ドメインに関する発話を行ったが、音声認識誤りによってドメイン選択誤りが生じ、ホテルドメインに遷移してしまった。U3で再び寺社案内ドメインに関する発話を行ったが、同様の音声認識誤りが生じた。このような場合、U3に対して選択すべきドメインは、前ターンのドメインでもなければ、音声認識結果を受理できるドメインでもない。このことが判定できれば、誤ったドメインで応答を続けることを防止する「S3正」のような応答を行うことができる。さらにU1の状態まで戻るかどうかをユーザに確認することで正しいドメインから対話を再開することが可能となる。また、この際ホテルドメインへの遷移を禁止することで再び誤ったドメイン遷移が起きることを防ぐことができる。

#### 4. 対話履歴を利用したドメイン選択

本稿では、ユーザ発話や対話履歴から得られる特徴量を入力とし、(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) それ以外のいずれかのドメイン、を選択する判別器を、対話データから学習する。以下ではまず、ドメイン選択のために利用する特徴量について説明する。

##### 4.1 ドメイン選択で利用する特徴量

分散型アーキテクチャの特徴であるドメインの追加・変更の容易性を確保するためには、ドメイン選択に利用する特徴量はドメインに依存しないものである必要がある。本稿ではドメイン選択を図2のように捉えており、これ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する。

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。利用した特徴量を図5に示した。P1からP9の各特徴量は、そ

P1: そのドメインに遷移してから、ユーザの肯定応答があった回数  
 P2: そのドメインに遷移してから、ユーザの否定応答があった回数  
 P3: そのドメインに遷移する前に、同じドメインでタスク達成（データベース検索の場合、情報提示があったか）されたことがあるか。  
 P4: そのドメインに遷移する前に、同じドメインであったことがあるか。  
 P5: そのドメインに遷移してから現在までに変化したスロット数  
 P6: そのドメインに遷移してから現在までのターン数  
 P7: スロットの変化の割合（ $=P6/P7$ ）  
 P8: システムからの質問への応答における否定応答の割合（ $=P3/(P2+P3)$ ）  
 P9: 対話におけるユーザの否定応答の割合（ $=P3/P7$ ）  
 P10: タスクの状態

図5 ひとつ前の応答を行ったドメインに関する特徴量

U1: ひとつ前の応答を行ったドメインで言語理解できた音声認識結果の音響スコア  
 U2: ひとつ前の応答を行ったドメインで言語理解できた音声認識結果の事後確率  
 U3: ひとつ前の応答を行ったドメインで言語理解できた音声認識結果の単語信頼度相加平均  
 U4: 音声認識結果に対する尤度最大のドメインが受理した音声認識結果の音響スコア  
 U5: 音声認識結果に対する尤度最大のドメインが受理した音声認識結果の事後確率  
 U6: 音声認識結果に対する尤度最大のドメインが受理した音声認識結果の単語信頼度相加平均  
 U7: 音響スコア（対数尤度）の差（ $=U1-U4$ ）  
 U8: 事後確率の比（ $=U2/U5$ ）  
 U9: 単語信頼度相加平均の比（ $=U3/U6$ ）

図6 ユーザ発話音声認識結果に関する特徴量

のドメインの推定が信頼できるかの観点から定義した。肯定応答が多い場合や、スロットの変化がある場合はそのドメインが信頼でき、逆に否定応答が多い場合などはそのドメインの信頼性は低いと考えられる。また、ドメインが信頼できる状況でもタスクの状態によってドメインの遷移しやすさが変化するため、これを特徴量としてタスクの種類に応じて用意した（P10）。文献13)のタスク分類に従ったときのスロットフィリング型タスクの場合は、埋まっていないスロットがある状態（タスク途中）と必要な全てのスロットが埋まっている状態（タスク達成）の2種類とした。また、データベース検索タスクの場合は、文献14)で定義した2状態「検索条件の指定」と「情報の提示要求」とした。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と同種類のものである。これを図6に示す。ひとつ前の応答を行ったドメインで解釈した認識結果に関するもの（U1~U3）と、音声認識結果に対する尤度最大のドメインで解釈した認識結果に関するもの（U4~U6）、及びそれらの比較（U7~

C1: ひとつ前の応答を行ったドメインを選択した場合の、そのドメインのタスクの状態  
 C2: ひとつ前の応答を行ったドメインで言語理解した場合、肯定応答かどうか  
 C3: ひとつ前の応答を行ったドメインで言語理解した場合、否定応答かどうか  
 C4: ひとつ前の応答を行ったドメインで言語理解した場合、変化するスロット数  
 C5: ひとつ前の応答を行ったドメインで言語理解した場合、変化する共有スロット数  
 C6: 音声認識結果に対する尤度最大のドメインを選択した場合の、そのドメインのタスクの状態  
 C7: 音声認識結果に対する尤度最大のドメインで言語理解した結果が、肯定応答かどうか  
 C8: 音声認識結果に対する尤度最大のドメインで言語理解した結果が、否定応答かどうか  
 C9: 音声認識結果に対する尤度最大のドメインを選択した場合に変化するスロット数  
 C10: 音声認識結果に対する尤度最大のドメインを選択した場合に変化する共有スロット（地名）数  
 C11: 音声認識結果に対する尤度最大のドメインが、それまでにタスク達成（データベース検索の場合、情報の取得）されたことがあるか  
 C12: 音声認識結果に対する尤度最大のドメインが、それまでに存在したか  
 C13: 音声認識結果を受理できた尤度最大のドメインを選択した場合、共有スロット（地名）が変更されるか

図7 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

U9)に分かれている。

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上でシステム本体に返す情報に相当する。これを図7に示す。このうち「ひとつ前の応答を行ったドメイン」を選択した場合にどのような状態になるかを表現したものがC1~C5、「音声認識結果に対する尤度最大のドメイン」を選択した場合にどのような状態になるかを表現したものがC6~C13である。ひとつ前の応答を行ったドメインを選択した場合の特徴量が少ないのは、先にひとつ前の応答を行ったドメインに関する特徴量を用意しているためである。

#### 4.2 ドメイン選択器の構成

学習に用いる対話データは、Woz法を用いて収集するのではなく、実際に音声対話システムを用いて収集する。これは、Woz法で対話データを収集すると、対話データ中にドメイン選択誤りが存在しないため、ドメイン選択誤りが生じた際の頑健性を持たない判別器が生成されるためである。

学習に用いる正解ラベルの付与は、各ターンごとにシステムの選択したドメインと、ユーザ発話の書き起しを基に以下に従って行う。

ラベル (I): ユーザ発話に対する正解ドメインが、ひとつ前の応答を行ったドメインと同じ場合

ドメイン	タスクの種類	音声認識の語彙サイズ	スロット数	記述言語
レストラン	データベース検索	1,562	10	Java
ホテル	データベース検索	741	9	Java
観光	データベース検索	1,573	4	Java
天気案内	スロットフィリング	87	3	Java
バス運行情報	スロットフィリング	1,621	3	Perl
全体	-	7,373	-	-

ラベル (II) : (I) 以外の場合で、ユーザ発話に対する正解ドメインが、N-best 音声認識結果の中で最も認識スコアの高い音声認識結果を解釈できたドメインである場合

ラベル (III) : 上記以外の場合

最尤の音声認識結果が複数のドメインで言語理解できた場合には、(II) に該当するドメインは複数存在する。この場合は、その各ドメインに対して判別器の算出した判別の信頼度を基にスコア付けを行い、ドメインを決定する。

## 5. 評価実験

### 5.1 評価用システムの構成

提案手法を評価するために、マルチドメイン音声対話システムを実装した。エキスパートを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、バス運行情報案内の5つである。それぞれの詳細を表1に示す。システム本体はJavaを用いて実装しているが、各エキスパートはどの言語でも動作するように設計している。レストラン、ホテル、観光、天気ドメインがJavaで実装されているのに対し、バスドメインはPerlによって実装されている。また「地名」属性のスロットは今回用意した5ドメインのいずれにも含まれていないため、共通の履歴を参照するように実装した。

音声認識エンジンはJulian<sup>15)</sup>を用いた。音声認識用文法は、各ドメインの言語理解部で用いた言語理解用文法から自動生成することにより得た。また、音響モデルはJulius ディクテーションキット付属<sup>15)</sup>の3000状態不特定話者PTMトライフォンモデルを利用した。

### 5.2 対話データの収集

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分ほど練習を行う。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う。同様の条件で3対話を行った。

データ収集時のシステム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。ここでは、10-best 音声認識結果のうち最も音響尤度の高い認識結果を言語理解できたドメインを選択した。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40加算して比

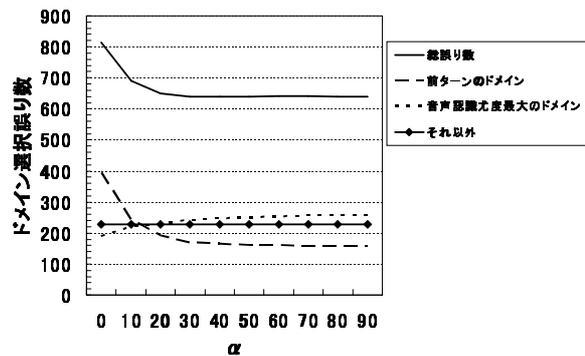


図8 従来手法によるドメイン選択の精度

較した。この値は予備実験に基づき決定した。

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。単語正解率が低いのは、文法外、語彙外発話による音声認識誤りが起きたときに、ユーザが同様の発話を繰り返して誤りを増加させる傾向があることに起因する。また、音声認識率の低い話者ほどタスク達成までの発話数が多くなるため、全体として低い単語正解率となっている。

### 5.3 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。

提案手法：提案するドメイン選択を行う。今回、ドメイン選択器をC5.0<sup>16)</sup>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、 $\alpha$ の値を0から90まで変化させて評価を行った際のドメイン選択誤りの数を正解ラベルごとにまとめた(図8)。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。 $\alpha=0$ のときは、ドメインを維持する制約を全く用いず、音声認識結果のみからドメイン選択を行う手法に相当する。図8に示すように、 $\alpha$ を大きくするほど、正解ラベルが「ひとつ前の応答を行ったド

メイン」であるものの誤り数が減少する。これは、ひとつ前の応答を行ったドメインを維持する制約によって、音声認識誤りによる誤ったドメイン遷移が抑制されるためである。 $\alpha$  が 40 以上では、ひとつ前の応答を行ったドメインで受理できる音声認識結果が、10-best 結果中に現われなくなるため誤り数は減少しない。逆に、 $\alpha$  を大きくするほど、正解ラベルが「音声認識尤度最大のドメイン」であるものの誤りが増加する。これは、ひとつ前の応答を行ったドメインが誤っていた場合、もしくは、ユーザがドメイン変更を行った場合に、ひとつ前の応答を行ったドメインを維持する制約によって正しいドメインが選択されなくなっていることを示す。 $\alpha = 45$  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。この中には、正解ラベルが「ひとつ前の応答を行ったドメイン」「音声認識を解釈できる最尤のドメイン」のどちらでもないものが 228 個含まれる。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。決定木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 2 に示す。改善が大きかったのは、正解ラベルが「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。この改善された 83 発話のうち 39 発話に関しては、ひとつ前に選択されていたドメインに戻すことで正しいドメインに戻すことができるものであった。このことから、「その他のいずれかのドメイン」であることが判定された時に、ドメインをひとつ前の段階に戻すかどうかを確認することが有効であると考えられる。残りの 44 発話に関しても、ユーザに否定されたドメインへの遷移を抑制して対話を再開することで、誤ったドメイン遷移が繰り返されることを防ぐことができる。

生成された決定木のうち最も上位に現れた特徴量は音声認識事後確率の比 (U8) であった。また、ひとつ前の応答を行ったドメインでの肯定の回数 (P2) や対話における否定の割合 (P10)、そのドメインに遷移してからのターン数 (P7) や変化したスロット数 (P6) など、ひとつ前の応答を行ったドメインに関する特徴の多くが上位に現れていた。その他、音声認識結果に対する尤度最大のドメインが、それまでのドメイン系列に存在するか (C12) やその時の解釈結果が否定応答であるか (C3) といった特徴が上位に現れていた。

表 2 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

## 6. ま と め

本稿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、本稿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。今後、今回の実験では改善が十分でなかった部分の検討を行う。また、ドメイン数の異なるシステムを用いた実験により、手法のドメイン拡張性を評価する予定である。

## 参 考 文 献

- 1) 駒谷和範, 他. 柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム. 電子情報通信学会技術研究報告, SP2001-113, 2001.
- 2) 音声ポケロケ. <http://www.lang.astem.or.jp/bus>.
- 3) T.Isobe et al. A study on domain recognition of spoken dialogue systems. In *Proc. EUROSPEECH*, 2003.
- 4) 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. 情報処理学会研究報告, 2003-SLP-45-20, 2003.
- 5) I.Lane et al. Topic classification and verification modeling for out-of-domain utterance detection. In *Proc. ICSLP*, pp. 2197–2200, 2004.
- 6) 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, Sept. 2001.
- 7) 長森誠, 他. マルチドメイン音声対話システムの構築手法. 情報処理学会研究報告, 2000-SLP-31-7, 2000.
- 8) 河口信夫, 他. 複数の音声対話システムの統合制御機構とその評価. 情報処理学会研究報告, 2001-SLP-36-10, 2001.
- 9) I.O'Neill et al. Cross domain dialogue modelling: An object-based approach. In *Proc. ICSLP*, Vol.I, 2004.
- 10) B.Pakucs. Towards dynamic multi-domain dialogue processing. In *Eurospeech 2003*, pp. 741–744, 2003.
- 11) 宮崎昇, 他. 音声対話システムの半自動統合による複数ドメイン対応. 日本音響学会秋季講演論文集, pp. 189–190, 2005.
- 12) 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.
- 13) M.Araki et al. A dialogue library for task-oriented spoken dialogue systems. In *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pp. 1–7, 1999.
- 14) K.Komatani et al. Contextual constraints based on dialogue models in database search task for spoken dialogue systems. In *Proc. EUROSPEECH*, pp. 877–880, Sep. 2005.
- 15) 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.
- 16) C5.0. <http://rulequest.com/index.html>.