マルチ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

神田直之[†] 駒谷和範[†] 中野幹生[††]
中臺一博[††] 辻野広司[††]
尾形哲也[†] 奥乃博[†]

複数のドメインを扱う音声対話システムにおいて，対話履歴から得られる特徴量を導入してより精度良くドメイン選択を行う手法を開発した．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) 1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) それ以外のドメイン，のいずれかという判別問題ととらえる．対話履歴から得られる特徴量を用いて上記を判別する決定木を，ドメイン選択の正解を与えた対話データから学習し，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実装し，これを用いて10名の被験者から対話データを収集した．この対話データを用いた評価実験の結果，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16.2%削減されることを確認した．

Robust Domain Selection Using Dialogue History in Multi-domain Spoken Dialogue Systems

Naoyuki Kanda,[†] Kazunori Komatani,[†] Mikio Nakano,[††]
Kazuhiro Nakadai,[††] Hiroshi Tsujino,[††] Tetsuya Ogata[†]
and Hiroshi G. Okuno[†]

We have developed a robust domain selection method using dialogue history in multi-domain spoken dialogue systems. We define domain selection as a 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 data. We implemented a multi-domain spoken dialogue system with 5 domains, and collected dialogue data from 10 subjects. The experimental result showed our method reduced 16.2% of domain selection errors, compared with a conventional method using speech recognition likelihoods only.

1. はじめに

これまで，飛行機の予約やバス運行情報案内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[5),10)]．これらのシステムの多くは，扱えるドメインが飛行機予約のみ，バス運行情報のみと1種類に限られており，ユーザの多様な要求に十分応えられるものではなかった．そこで本研究では，シングルドメインの音声対話コンポーネントの統合により，複数のドメインを扱えるシステム（**マルチドメイン音声対話システム**）を開発した．1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，複雑なタスクも扱える．またドメイン間での履歴の共有により，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，まずユーザ要求の対象ドメインを推定する**ドメイン選択**処理が不可欠である．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」の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．しかしこの方法は，対話が冗長で自然さに欠けるうえに，ユーザはシステム設計者の定義したドメインの区切りを理解することを強制され，ユーザ

† 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
†† 株式会社ホンダ・リサーチ・インスティチュート・ジャパン
Honda Research Institute Japan Co., Ltd.

に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の応答すべきドメインの推定が行われてきた[3),4),11)]．これらの研究では，1 発話に対する音声認識結果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．このため，音声認識結果が誤りである場合にはドメイン選択も誤りとなることが多い．

音声認識誤りが生じた場合でも正しいドメインで対話を行うには，それまでの対話の流れも考慮したドメイン選択が必要である．本研究では対話履歴から得られる特徴量を導入することで，音声認識誤りに対してより頑健なドメイン選択を行う．

また，マルチドメイン音声対話システムでは，シングルドメインのシステムよりも構築にかかる労力が大きい．したがって，新たなドメインの追加や，既存のドメインの改変の容易さ（ドメインの保守性・拡張性）が高いシステムアーキテクチャが望ましい．これはドメイン選択手法においても同様であり，ドメインの変更や追加にも対応できる枠組みが求められる．

本研究では，マルチドメイン音声対話システムにおいて，対話履歴から得られる特徴量を導入し，音声認識誤りに頑健なドメイン選択を行う手法を提案する．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) 1 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) それ以外のドメイン，のどれに該当するかを，対話履歴とユーザ発話から判別する問題ととらえる．対話履歴から得られる特徴量を導入して判別器を構成することにより，音声認識誤りに対してより頑健なドメイン選択が可能となる．また，上記の 3 クラスは，ドメインが増減しても一様に定義できるため，保守性・拡張性が高い．

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，シングル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て設計された場合，あるドメインの改変の影響がシステム全体に及ぶため，ドメイン内部での改変や，新たなドメインの追加が困難となる．そこで，各ドメインが独立に設計できる分散型[6)] のアーキテクチャが提案されている[6),8),9),11),12),14),16)]．このアーキテクチャでは，ドメインごとに独立して記述できる部分と，ドメイン間の依存関係の考慮が必要な部分を切り分ける．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でき，各ドメインの改良や追加が容易となる．

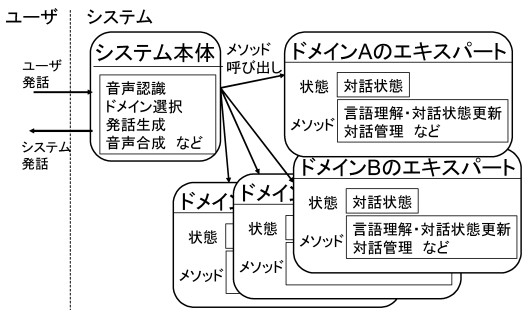

図 1 マルチドメイン音声対話システムにおける分散型のシステムアーキテクチャ
Fig. 1 Distributed architecture for multi-domain spoken dialogue systems.

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[7)] に基づきシステムを設計した．システムは大きく分けて，各ドメインでの対話を担当するエキスパートと，それらを統括するシステム本体に分かれる（**図 1**）．ユーザ発話の処理（言語理解や応答生成など）は各ドメインごとにエキスパートが行い，システム本体ではどのエキスパートにユーザ発話の処理を任せるかの振り分けを行う．また，特定のスロットに関しては，一定のプロトコルに従い，システム本体を仲介役としてドメイン間で共有を行う．各ドメインでの処理は各エキスパート内で完結しているため，システム本体とのプロトコルに従えば，システム設計者は個々のドメインを独立に設計できる．これにより各ドメインの改良や追加が容易となる．

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を担当する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という判定が重要となる．本研究ではこれを**ドメイン選択**と呼ぶ．分散型アーキテクチャの特長である，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もドメイン非依存である必要がある．

3. ドメイン選択の課題

分散型アーキテクチャの枠組みのもとでは，ドメインの保守性・拡張性を備えたドメイン選択の設計が必要である．さらに，ドメイン選択は音声認識誤りに対して頑健でなければならない．

これまでの多くの手法では，音声認識結果から各ドメインごとにスコアを与え，そのスコアが最も高いドメインを選択する[3),4),11)]．しかし，ドメインへのスコアの算出方法が音声認識結果のみに依存しているため，音声認識結果が誤りである場合，ドメイン選択も誤り

となることが多い．さらに，新たなドメインを追加する場合にそのドメインの精密な言語モデルを必要とするため，ドメインの拡張性が低い．

また，ドメイン選択の際に 1 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)，16) では 1 つ前の応答を行ったドメインに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 8) では，サブゴールが達成されるまでドメイン遷移を行わない．これらの手法では，1 つ前の応答を行ったドメインが正しい場合，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，1 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．この問題は，1 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．本研究では，1 つ前のドメイン推定の信頼性を表現する情報を，ドメインの履歴や状態から得ることにより，この問題を解決する．

4. ロバストで拡張性を備えたドメイン選択

4.1 本研究でのドメイン選択の定義

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) 1 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) それ以外のドメイン，のいずれであるかを選択する問題ととらえる（**図 2**）．ドメイン選択器は，これらの判別器を対話データから学習して構成する．選択肢 (I) は，1 つ前の応答を行ったドメインを維持する場合に相当し，選択肢 (II) は，ユーザの要求するドメインが変化したときに，音声認識結果に基づきドメイン選択を行う場合に相当する．このように，この枠組みは従来研究でのドメイン選択を包含する．さらに選択肢 (III) を定義し判別することで，選択肢 (I) や (II) が有効でない場合にも対処する．

音声認識の成否とドメイン選択を分けて考えることで，音声認識結果が誤りである場合でも，1 つ前に応答したドメインが正しければ，それを維持できる．この例を**図 3** に示す．ユーザはまず U1 でレストランに関する発話を行う．次に U2 でも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われるべきであるが，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，選択肢 (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．

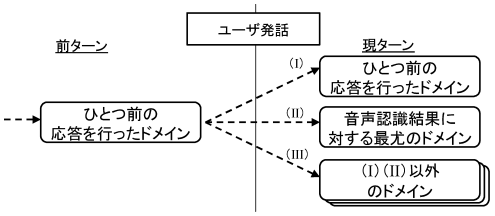

図 2 ドメイン選択の概略
Fig. 2 Overview of our domain selection.

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（寺社ドメイン）」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．（寺社ドメイン）
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 選択肢 (I) の場合の対話例
Fig. 3 Example in which choice (I) should be selected.

さらに選択肢 (III) を定義し判別することで，通常の制御が適切でない場合，つまり，1 つ前の応答を行ったドメインや音声認識結果に基づくドメインのエキスパートに制御権を与えるのが適当でない場合を検出できる．これは誤りが連続した場合に相当するが，ドメインは正しいが音声認識が誤りであるのか，ドメインそのものが誤りであるのかの判断が必要な点で，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る．この例を**図 4** に示す．ユーザは U1 でホテルドメインに関する発話を行おうとしたが，いい淀んだため音声認識誤りが生じ，レストランドメインに遷移してしまった．U2 で再びホテルドメインに関する発話を行ったが，さらに音声認識誤りが生じた．このような場合，U2 に対して選択すべきドメインは，1 つ前の応答を行ったドメイン（レストラン）でもなければ，音声認識結果に対するドメイン（バス）でもない．このような状況を検出することで，誤ったドメインでの対話の継続を防止できる．たとえば，図 4 の S2 正のように，「［選択肢 (I) のドメイン］または［選択肢 (II) のドメイン］の情報についてお尋ねですか？」といった確認が生成できる．

これらの選択肢は，個々のドメインの判別ではなく，

☆ 選択肢 (I) と (II) が同じドメインとなる場合は (I) とする．

U1：**京都 •••**（ホテルドメイン）
（「京都料理（レストランドメイン）」と誤認識）
S1：京都料理でレストランを検索します．50 件あります．
U2：**いや京都グランドホテルの場所**（ホテルドメイン）
（「京都グランドホテルでバス（バスドメイン）」と誤認識）
↓
S2 誤：京都グランドホテル前からバスでどこまで行かれますか？
S2 正：レストラン又はバスの情報についてお尋ねですか？

図 4 選択肢 (Ⅲ) の場合の対話例
Fig. 4 Example in which choice (Ⅲ) should be selected.

ドメイン間の時系列上での相対的な関係を表すため，ドメイン数が増減しても一様に定義できる．そのため，得られた判別器はドメインの数に依存せず利用でき，保守性・拡張性が高いドメイン選択を実現できる．

4.2 対話履歴を利用したドメイン選択

前節で定義したドメイン選択を行うために用いた特徴量について説明する．本研究では，従来用いられてきた，音声認識結果から得られる情報に加えて，対話履歴から得られる特徴量の導入により，ドメイン選択の各選択肢が妥当かどうかを表現した．ここでは，**選択肢 (I) の妥当性を表す特徴量（図 5）**と，**選択肢 (Ⅱ) の妥当性を表す特徴量（図 6）**を定義し利用した．選択肢 (Ⅲ) は選択肢 (I) と (Ⅱ) の補集合であるため，選択肢 (I) と (Ⅱ) のいずれでもない場合に選択される．特徴は全部で 32 個用意した[17] が，ここでは 4.3 節で有効とされたもののみを列挙している．

選択肢 (I) は，1 つ前の応答を行ったドメインが信頼でき，かつドメインを維持すべき場合に選択される．ここで利用した特徴量を図 5 に示す．本研究では，対話履歴を用いて 1 つ前の応答を行ったドメインが信頼できるかを表現する．たとえば，あるドメインでの対話中にユーザの肯定応答（「はい」「そうです」など）が多ければ，そのドメインが正しい可能性は高いとする（I_1）．反対に，あるドメインでの対話中にシステムのスロットの状態に変化が見られなければ，ユーザ要求がそのドメインで処理されておらず，そのドメインが誤りである可能性が高い（I_5，I_7）．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（I_{10}）．スロットフィリング型タスク[1] 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）と必要なすべ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[1] は，文献 18) 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．そのほかに，そのドメインを選択した場合の履歴・状態も指標とした．たとえば，そのドメインを選択したときに変化するスロット数が少ない場合は，そのドメインで音声認識結果をうまく処理できないと考えられるため，そのドメイン選択は誤りである可能性が高いとする（I_{13}）．また，受理された音声認識結果の事後確率を，音響的・言語的にそのドメインが信頼できるかの指標として利用した（I_{14}）．これは従来の研究で用いられていた指標と同様のものである．

I_1: (I) 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
I_2: (I) 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
I_3: (I) 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，情報の提示があったか）されたことがあるか
I_4: (I) 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
I_5: (I) 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
I_6: (I) のドメインに遷移してから現在までのターン数
I_7: スロットの変化の度合（$= I_5/I_6$）
I_8: システムからの質問への応答における否定応答の割合（$= I_2/(I_1 + I_2)$）
I_9: 対話におけるユーザの否定応答の割合 ($= I_2/I_6$)
I_{10}: (I) のドメインの，現在のタスクの状態
I_{11}: (I) を選択した場合の，そのドメインのタスクの状態
I_{12}: (I) のドメインで言語理解した場合，否定応答となるかどうか
I_{13}: (I) を選択した場合，変化するスロット数
I_{14}: (I) のドメインで言語理解できた音声認識結果の文としての事後確率

図 5 選択肢 (I) の妥当性を表す特徴量
Fig. 5 Features supporting choice (I).

次に，選択肢 (Ⅱ) の妥当性を表す特徴量を図 6 に示す．ここでも，対話履歴から得られる特徴量を利用した．ドメイン選択後のタスクの状態や，変化するスロット数などはここでも利用した（II_1，II_3）．また，選択肢 (Ⅱ) をとることで急激な話題の変化が生じる場合は，そのドメイン選択が誤りである可能性が高い．ドメインが (Ⅱ) に変化し，かつドメイン間で共有しているスロット値が変化する場合，ドメインが変化するだけの場合や共有スロットが変化するだけの場合と比べて話題の変化が大きいため，そのドメイン選択が音声

II_1: (II) を選択した場合の，そのドメインのタスクの状態
II_2: (II) のドメインで言語理解した結果が，否定応答かどうか
II_3: (II) を選択した場合に変化するスロット数
II_4: (II) を選択した場合に変化する共有スロット数
II_5: (II) のドメインが，それまでに存在したか
II_6: (II) のドメインが受理した音声認識結果の文としての事後確率
II_7: (II) のドメインが受理した音声認識結果に含まれる単語の信頼度の相加平均
II_8: (I) と (II) のドメインで各々言語理解できた音声認識結果の音響スコアの差
II_9: (I) と (II) のドメインで各々言語理解できた音声認識結果に含まれる単語の信頼度の相加平均の比

図 6 選択肢 (II) の妥当性を表す特徴量
Fig. 6 Features supporting choice (II).

認識誤りによるものである可能性が高いとする（II_4）．そのほかに，音声認識尤度最大の解釈結果を持つドメインで受理された，音声認識結果の事後確率や単語信頼度の相加平均を，選択肢 (II) の音響・言語的な尤もらしさの指標とする（II_6，II_7）．また，選択肢 (I) で受理された音声認識結果との比較を行い，その音声認識結果が曖昧性のあるものかどうかの指標として利用する（II_8, II_9）．

4.3 ドメイン選択で利用する特徴量の選択

本研究では，4.2 節で定義した各特徴量を用いてドメイン選択器を構成した．特徴量の中には判別に悪影響を及ぼすものがあるため，事前に以下の手順により特徴量選択を行った．以下では，ドメイン選択器の学習用対話データと，ドメイン選択精度の評価用対話データが用意されていると仮定している．ただし，今回は対話データが少ないため，クロスバリデーション法を用いて評価を行った．

(1) 文献 17) で定義した特徴量の集合を $\boldsymbol{F}$ とする．
(2) $\boldsymbol{F}$ に含まれる特徴量から 1 つを選択し，(3)〜(4) を行う．これを $\boldsymbol{F}$ に含まれる特徴量すべてに対して行う．
(3) 選択された特徴量を a とし，$\boldsymbol{F}$ から a を取り除きドメイン選択器を学習する．
(4) 得られたドメイン選択器を用いてドメイン選択精度を算出する．
(5) 特徴量を取り除くことでドメイン選択の精度が向上した（もしくは変化がなかった）場合，最も精度の向上が大きかった特徴量を $\boldsymbol{F}$ から取り除き，(2) に戻る．
(6) どの特徴量を除いてもドメイン選択精度が悪化するようになった時点で終了する．このときの $\boldsymbol{F}$ が判別に利用する特徴量集合となる．

5. 評価実験

5.1 評価用データの収集

提案手法の評価用データを収集するため，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[10] 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．実際，レストラン，ホテル，寺社，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 で実装されている．また本システムは，複数のドメインにまたがるスロットの内容を共有する機能を有している．これにより，あるドメインで話題になったスロットの値を，後続するドメインに引き継がせることができる．今回の 5 ドメインでは地名属性に対応するスロットを共有させた☆．

音声認識エンジンは Julian[15] 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．音響モデルには Julius ディクテーションキット付属[15] の 3000 状態 PTM トライフォンモデルを利用した．またシステムからの応答には，図 3，図 4 の対話例に示されるように音声認識結果を含ませることで暗黙的な確認を行った．さらにシステム発話は，音声合成を行うとともに画面上にテキストで表示した．これらによりシステム・ユーザ間に誤解が生じないようにした．

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．

データ収集時のシステムでは，音声認識結果の 10-best 解の上位から順に言語理解を行い，最初に言語理解できた発話から得られたドメインを選択した．ただし，1 つ前の応答を行ったドメインには，音響尤度に 40 を加算して比較した．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 2,205 発話（221 発

☆ 地名属性は，ドメインにより表現や粒度が異なるので，緯度・経度情報として共有することで，各ドメインにおける表現に変換できるよう実装した．

表 1 各ドメインの概要
Table 1 Specifications of each domain.

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
レストラン	DB	1,562	10
ホテル	DB	741	9
寺社	DB	1,573	4
天気案内	SF	87	3
バス運行情報	SF	1,621	3
全体	—	7,373	—

DB：データベース型検索タスク
SF：スロットフィリング型タスク

話/人，74 発話/対話）で，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．収集したデータには，ユーザ発話の音声認識結果が肯定応答（「はい」など）であったものが 274 発話含まれる．これらに対しては，正解がほぼ「(I) 1 つ前の応答を行ったドメイン」となるため，以下の評価はこれらを除いた 1,931 発話を対象とした☆．なお肯定発話を含む全 2,205 発話に対する評価も併記している．

収集したユーザ発話には，各ターンごとにシステムが応答を行ったドメインを記録し，ユーザ発話の書き起こしをもとに以下に従って正解ラベルを付与した．

(1) ユーザ発話に対する正解ドメインが，1 つ前の応答を行ったドメインと同じ場合，ラベル (I) を付与する．
(2) (I) 以外の場合で，ユーザ発話に対する正解ドメインが，音声認識結果の N-best 解の中で最も認識スコアの高い結果を解釈できたドメインである場合，ラベル (II) とする．
(3) 上記以外の場合にはラベル (III) を付与する．

今回の対話データに対して学習器 C5.0[2] を用いて生成した決定木を**図 7** に示す．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スコアの差（II_8）であった．また，1 つ前の応答を行ったドメインでの肯定の回数（I_1）や対話における否定の割合（I_9），そのドメインに遷移してから変化したスロット数（I_5）など，対話履歴から得られる特徴の多くが上位に現れた．

この決定木を用いてドメイン選択を行った場合のシステム動作について，**図 8** の対話を例にとり説明する．ここではユーザは U1 で「明日の気温をお願いします」と発話したが，「パスタをお願いします」と誤認識され，レストランドメインへ遷移してしまった．このとき，「パスタ」の単語信頼度が低かったため，S1 のようにシステムからユーザへの確認が行われた[13]．これに対しユーザは，「天気をお願いします」と発話したが（U2），今度は「円福寺をお願いします」と認識されてしまった．このとき，選択肢 (I) はレストランドメイン，選択肢 (II) は寺社ドメインに対応する．(II) のドメインと (I) のドメインで各々受理した音声認識結果のスコアの差は 26.2 であった．また，(II) のドメインで受理した音声認識結果に含まれる単語の信頼度の相加平均は 0.64 と小さかった．(I) のドメインに遷移してからのユーザからの確認応答は 1 つもなく，そのドメインでタスクが達成されたこともなかった．さ

☆ 「いいえ」などの否定応答は，選択肢 (I) と (III) を判定する必要があるため評価対象とした．

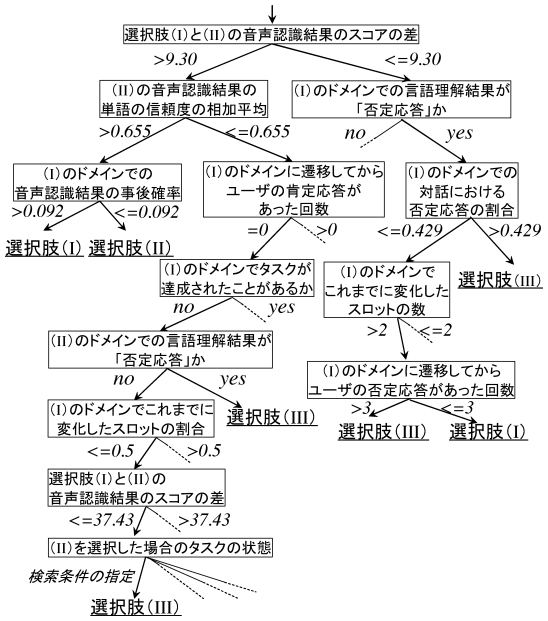

図 7 対話データから得られた決定木
Fig. 7 Decision tree constructed from dialogue data.

U1：**明日の気温をお願いします**（天気ドメイン）
（「パスタをお願いします（レストランドメイン）」と誤認識.）
S1：フードタイプがパスタのレストランを検索してもよろしいですか？（レストランドメイン）
U2：**天気をお願いします**（天気ドメイン）
（「円福寺をお願いします（寺社ドメイン）」と誤認識．ドメイン選択の結果，選択肢 *(III)* が選択された.）
S2：レストランまたは寺社についてお尋ねですか？

図 8 (III) その他のドメインが選択された対話例
Fig. 8 Example in which choice (III) was selected.

表 2 全発話に対するドメイン選択結果の confusion matrix（ベースライン手法/本手法）
Table 2 Confusion matrix in domain selection for all utterances (baseline / our method).

正解 \ 判別結果	(I)	(II)	(III)	計 (再現率)
(I)1 つ前のドメイン	1,289 / 1,291	162 / 85	0 / 75	1,451 (0.89 / 0.89)
(II) 音声認識結果に対する最尤のドメイン	84 / 99	299† / 256†	0 / 28	383 (0.74 / 0.62)
(III) その他のドメイン	293 / 172	78 / 42	0 / 157	371 (0 / 0.42)
計 (適合率)	1,666 / 1,562 (0.77) / (0.83)	539 / 383 (0.52) / (0.62)	0 / 260 (—) / (0.60)	2,205 (0.712 / 0.765)

† 音声認識結果に対する最尤のドメインが複数存在し，その曖昧性を解消できなかった誤りが，それぞれ 17 ずつ含まれる

表 3 肯定発話を除いた場合のドメイン選択結果の confusion matrix（ベースライン手法/本手法）
Table 3 Confusion matrix in domain selection after removing affirmative utterances (baseline / our method).

正解 \ 判別結果	(I)	(II)	(III)	計 (再現率)
(I)1 つ前のドメイン	1,031 / 1,023	162 / 87	0 / 83	1,193 (0.86 / 0.86)
(II) 音声認識結果に対する最尤のドメイン	78 / 95	299† / 247†	0 / 35	377 (0.75 / 0.62)
(III) その他のドメイン	283 / 161	78 / 42	0 / 158	361 (0 / 0.44)
計 (適合率)	1,392 / 1,279 (0.74) / (0.80)	539 / 376 (0.52) / (0.62)	0 / 276 (—) / (0.57)	1,931 (0.680 / 0.732)

† 音声認識結果に対する最尤のドメインが複数存在し，その曖昧性を解消できなかった誤りが，ベースライン手法で 17，本手法で 15 存在する

らに，変化したスロットはまだ 1 つもない☆ため，変化したスロットの対話における割合は 0 である．(II) を選択した場合のタスクの状態は「検索条件の指定」であった．これらの特徴量から，図 7 の決定木をたどると，選択肢 (III) 「その他のドメイン」が選択される．その結果，誤ったドメインで対話を続けることを防止するための,「レストランまたは寺社についてお尋ねですか？」という確認を行える（S2）．

5.2 ドメイン選択の精度の評価

以下の 2 種類の方法で，ドメイン選択を行った場合の誤り数を比較評価した．

ベースライン手法： 音声認識結果を受理できたドメインのうち，認識スコアが最大のものを選択する．ただし，1 つ前の応答を行ったドメインには，認識スコアに値 α を加算して比較した．

本手法： 提案するドメイン選択を行う．今回，ドメイン選択器は C5.0[2] 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 1 つのドメインを選択して正解判定を行った．

ベースライン手法で，α の値を 0 から 100 まで変化させた際のドメイン選択誤りの数を調査した．ここでは α の値が大きいほど，1 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まったく用いず，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．$\alpha = 35$ としたときに，誤り数が最も少なくなり，618 であった．ドメイン選択誤り率は 32.0%（=618/1,931）であった．この中には，正解ラベルが「1 つ前の応答を行ったドメイン」「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 361 個含まれる．これらはベースライン手法の枠組では正解を選択できない発話である．

次に，本手法による評価を示す．**表 2** に全データに対する場合，**表 3** に肯定発話を除いた場合のドメイン選択結果を，それぞれ confusion matrix の形で示す．表中の各項目では，左の数字がベースライン手法，右の数字が本手法によるドメイン選択結果を表す．また，対角線上の項目がドメインが正しく選択された発話数に相当し，それ以外の項目がドメイン選択を誤った発話数に相当する．ここでは被験者ごとにデータを分割した 10-fold のクロスバリデーションにより評価結果を得た．この評価結果は，決定木のカットオフパラメータを実験的に最適化した場合の値を示している．表 2 の結果の方が，判別の容易な肯定発話が含まれる分，表 3 と比べてドメイン選択精度は全体的に 3%程度高い．

☆ ユーザへ確認中のスロットは，変化したスロットとして計数していない．

以下，表 3 について詳しく説明する．まず，本手法によりドメイン選択誤りの数が 618 から 518 まで減少し，全体のドメイン選択精度が 68.0%から 73.2%へと 5.2%向上した．ドメイン選択誤りの削減率は 16.2%（100/618）である．また，ベースライン手法では選択肢 (Ⅲ) は選ばれない（表 3 の 3 列目）のに対し，本手法では 158 発話を検出している．これは，音声認識結果に対するドメインと 1 つ前に応答したドメインの両方が誤りである状況の約半分において，回復戦略をとるべき状況を検出できたことを示している．さらに，選択肢 (I) では，適合率が 0.74 から 0.80 に向上し，F 値が 0.80 から 0.83 に上昇した．また，選択肢 (Ⅱ) では，再現率が 0.75 から 0.62 へと減少したが，適合率が 0.52 から 0.62 へと向上し，F 値は 0.61 で変わらなかった．このように本手法では，それまでのドメイン選択精度を落とすことなく，従来検出できなかった選択肢 (Ⅲ) を検出している．

今回の評価実験では，本手法によるドメイン選択精度が 73.2%であり，いまだ 26.8%のドメイン選択誤りが残っている．このうち，選択肢 (Ⅱ) の判別に関しては，主に音声認識の精度がドメイン選択の精度に影響を与えている．今回の実験では音声認識の単語正解率が 63.3%と低い中での評価であったために，選択肢 (Ⅱ) の判別精度が低くなっていると考えられる．また，実際は選択肢 (I) もしくは (Ⅱ) であるものを，選択肢 (Ⅲ) と誤って判定したものが，118 発話存在した．ここで，選択肢 (Ⅲ) であるものを選択肢 (I) や (Ⅱ) と誤って選択した場合には，誤ったドメインでそのまま対話が進行してしまうのに対し，本来選択肢 (I) や (Ⅱ) であるものを選択肢 (Ⅲ) に誤っても，ユーザへのドメインの確認などのドメイン回復戦略が行われ，誤ったドメインでの対話が進行するわけではない．そのため，これらの誤りが対話に及ぼす悪影響は少ない．

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音声認識誤りに対してより頑健にドメインを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 16.2%削減されることを確認した．

本研究の意義を以下に示す．

- 1 つ前の応答を行ったドメインと音声認識結果に対する最尤のドメインの両方が適切でない状況を検出する必要性を指摘し，対話から得られる情報を用いることでその検出を可能とした．これまで，音声認識結果からドメインを推定するものや，1 つ前の応答を行ったドメインを信頼してそのドメインを維持する研究は行われていたが，音声認識結果から得られるドメインと 1 つ前のドメインのいずれもが妥当でない状況の存在を指摘した研究事例はない．本研究では，これらの状況を検出することで誤ったドメインでの対話の継続を防止できることを指摘し，ドメインの妥当性を表現する特徴量の導入によりその検出を行った．
- 様々な特徴量を用いた，ドメインの保守性・拡張性の高いドメイン選択法を実現した．これまでのドメイン選択では，各ドメインに対して判別スコアを算出していた．そのため，対話データなどから判別スコアを算出する場合には，ドメインを追加するたびに新たな対話データの収集が必要であるという問題があった．これに対し本手法では，ドメイン選択の対象を 3 クラスに抽象化することにより，ドメインの増減に対応できるドメイン選択法を実現した．

今後は，選択肢 (Ⅲ) が検出された際の最適なドメイン回復戦略の適用方法が課題となる．また今回の実験では改善が十分ではなかった部分に関して，新たな特徴量の設計などの検討が必要である．

謝辞 本研究の一部は，科学研究費補助金（基盤研究（A），特定領域「情報学」，若手研究（B）），21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」の支援を受けた．

参 考 文 献

1) Araki, M., Komatani, K., Hirata, T. and Doshita, S.: A Dialogue Library for Task-oriented Spoken Dialogue Systems, *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pp.1–7 (1999).
2) C5.0. http://rulequest.com/index.html
3) Isobe, T., Hayakawa, S., Murao, H., Mizutani, K., Takeda, K. and Itakura, F.: A Study on Domain Recognition of Spoken Dialogue Systems, *Proc. EUROSPEECH*, pp.1889–1892 (2003).
4) Lane, I.R., Kawahara, T., Matsui, T. and Nakamura, S.: Topic classification and verification modeling for out-of-domain utterance detection, *Proc. ICSLP*, pp.2197–2200 (2004).
5) Levin., E., Narayanan, S., Pieraccini, R., Biatov, K., Bocchieri, E., Fabbrizio, G.D., Eckert, W., Lee, S., Pokrovsky, A., Rahim, M., Ruscitti, P. and Walker, M.: The AT&T-DARPA communicator mixed-initiative spoken dialogue system, *Proc. ICSLP*, Vol.2, pp.122–

125 (2000).
6) Lin, B., Wang, H. and Lee, L.: A Distributed Agent Architecture for Intelligent Multi-Domain Spoken Dialogue Systems, *IEICE Trans. Information and Systems*, E84-D(9), pp.1217–1230 (2001).
7) Nakano, M., Hasegawa, Y., Torii, T., Takeuchi, Y., Nakadai, K., Tsujino, H., Kanda, N. and Okuno, H.G.: A Two-Layer Model for Behavior and Dialogue Planning in Conversational Service Robots, *Proc. IROS*, pp.1542–1548 (2005).
8) O'Neill, I., Hanna, P., Liu, X. and McTear, M.: Cross Domain Dialogue Modelling: An Object-Based Approach, *Proc. ICSLP*, Vol.I, pp.205–208 (2004).
9) Pakucs, B.: Towards Dynamic Multi-Domain Dialogue Processing, *Proc. Eurospeech*, pp.741–744 (2003).
10) 音声ポケロケ.
http://www.lang.astem.or.jp/bus
11) 安田宜仁，堂坂浩二，相川清明，上野晋一：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).
12) 宮崎　昇，甘粕哲郎，富久昭弘，萩野輝雄：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).
13) 駒谷和範，河原達也：音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理，情報処理学会論文誌，Vol.43, No.10, pp.3078–3086 (2002).
14) 河口信夫，長森　誠，松原茂樹，稲垣康善：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).
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) 長森　誠，河口信夫，松原茂樹，外山勝彦，稲垣康善：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).
17) 神田直之，駒谷和範，中野幹生，中臺一博，辻野広司，尾形哲也，奥乃　博：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，情報処理学会研究報告，2006-SLP60-11 (2006).
18) 神田直之，駒谷和範，尾形哲也，奥乃　博：データベース検索タスクにおける対話文脈を利用した音声言語理解，情報処理学会論文誌，Vol.47, No.6, pp.1802–1811 (2006).

(平成 18 年 3 月 31 日受付)
(平成 19 年 2 月 1 日採録)

神田　直之（正会員）

2004 年京都大学工学部情報学科卒業．2006 年同大学院情報学研究科知能情報学専攻修士課程修了．在学中は音声対話システムの研究に従事．現在，株式会社日立製作所勤務．

駒谷　和範（正会員）

1998 年京都大学工学部情報工学科卒業．2000 年同大学院情報学研究科知能情報学専攻修士課程修了．2002 年同大学院博士後期課程修了．京都大学博士（情報学）．同年より京都大学大学院情報学研究科助手．音声対話システムの研究に従事．情報処理学会平成 16 年度山下記念研究賞，FIT2002 ヤングリサーチャー賞受賞．電子情報通信学会，言語処理学会，人工知能学会，ACL 各会員．

中野　幹生（正会員）

1988 年東京大学教養学部基礎科学科第一卒業．1990 年同大学院理学系研究科相関理化学専攻修士課程修了．1990〜2004 年日本電信電話（株）にて，自然言語処理，音声対話システムの研究に従事．この間 2000〜2002 年 MIT 計算機科学研究所客員研究員．博士（理学）．2004 年より（株）ホンダ・リサーチ・インスティチュート・ジャパン，シニア・リサーチャ．音声コミュニケーションの研究に従事．言語処理学会，人工知能学会，日本ロボット学会，電子情報通信学会，ACL，ACM，IEEE 各会員．

中臺 一博

1970 年生．1993 年東京大学工学部電気工学科卒業．1995 年同大学院工学系研究科情報工学専攻修了．同年日本電信電話株式会社入社，1997 年 NTT コムウェア（株）出向後，1999 年退職．同年，JST ERATO 北野共生システムプロジェクト研究員．2003 年 5 月より（株）ホンダ・リサーチ・インスティチュート・ジャパン，シニアリサーチャ．博士（工学）．2006 年 4 月より，東京工業大学大学院情報理工学研究科客員助教授兼務．主にロボット聴覚，実時間情報統合，音環境理解の研究に従事．IROS 2001 BEST Paper Nomination Finalist，2002 年第 2 回船井情報科学振興賞等受賞．日本人工知能学会，日本音響学会，ヒューマンインタフェース学会，IEEE 各会員．

辻野 広司

1984 年東京工業大学理学部情報科学科卒業．1986 年同大学院情報科学専攻修士課程修了．1987 年（株）本田技術研究所入社．2003 年より（株）ホンダ・リサーチ・インスティチュート・ジャパン，チーフ・リサーチャ．脳型コンピュータ，知能システム，ヒューマンロボットインタフェース，画像認識等の研究に従事．IEEE，SFN，INNS，日本ロボット学会，人工知能学会，日本ソフトウェア科学会各会員．

尾形 哲也（正会員）

1993 年早稲田大学理工学部機械工学科卒業．日本学術振興会特別研究員，早稲田大学理工学部助手，理化学研究所脳科学総合研究センター研究員，京都大学大学院情報学研究科講師を経て，2005 年より同助教授．博士（工学）．この間，早稲田大学ヒューマノイド研究所客員助教授．人間とロボットのインタラクションと協調，神経回路モデル等の研究に従事．2000 年度日本機械学会論文賞，IEA/AIE-2005 最優秀論文賞等を受賞．RSJ，JSME，JSAI，IEEE 等各会員．

奥乃 博（正会員）

1972 年東京大学教養学部基礎科学科卒業．日本電信電話公社，NTT，JST，東京理科大学を経て，2001 年より京都大学大学院情報学研究科知能情報学専攻教授．博士（工学）．この間，スタンフォード大学客員研究員，東京大学工学部客員助教授．人工知能，音環境理解，ロボット聴覚，音楽情報処理の研究に従事．1990 年度人工知能学会論文賞，IEA/AIE-2001，2005 最優秀論文賞，IEEE/RSJ IROS-2001 Best Paper Nomination Finalist，第 2 回船井情報科学振興賞等受賞．JSAI，RSJ，ACM，IEEE 等各会員．本学会英文図書出版委員．