情報処理学会論文誌　Vol. 50　No. 2　488–500 (Feb. 2009)

マルチドメイン音声対話システムにおける トピック推定と対話履歴の統合によるドメイン選択手法

池田智志[†1]　駒谷和範[†1]
尾形哲也[†1]　奥乃博[†1]

本論文では，マルチドメイン音声対話システムにおいて，システム想定外発話に対しても頑健に，応答すべきドメインを決定する方法について述べる．想定外発話は言語理解誤りを引き起こし，ドメイン選択誤りの原因となる．そこで本論文ではまず，ユーザが意図したドメインをトピックとして定義し，Web から大量に収集した学習文書と，Latent Semantic Mapping を用いてトピックを推定する．次に，対話履歴とトピック推定を決定木を用いて統合し，想定外発話に頑健なドメイン選択器を構成した．対話履歴とトピック推定は相補的な情報であり，これらの統合により高精度なドメイン選択が実現できる．つまり我々が開発したトピック推定は，1 発話のみを対象とするが，想定外発話に頑健である．一方で対話履歴は，対話の流れを考慮できるが，想定外発話による悪影響を強く受ける．話者 10 名 2,191 発話を用いた評価実験により，従来手法と比較してドメイン選択誤りを 14.3%削減した．

Integrating Topic Estimation and Dialogue History for Domain Selection in Multi-Domain Spoken Dialogue Systems

Satoshi Ikeda,[†1] Kazunori Komatani,[†1]
Tetsuya Ogata[†1] and Hiroshi G. Okuno[†1]

We present a method of robust domain selection against out-of-grammar (OOG) utterances in multi-domain spoken dialogue systems. We first define a *topic* as a domain from which the user wants to retrieve information, and estimate it as the user's intention. This topic estimation is enabled by using a large amount of sentences collected from the Web and Latent Semantic Mapping (LSM). The topic estimation results are reliable even for OOG utterances. We then integrated both topic estimation results and dialogue history to construct a robust domain classifier against OOG utterances. The experimental results using 2,191 utterances showed that our integrated method reduced domain selection errors by 14.3%.

1. はじめに

電話などのインタフェースを通して，一般ユーザが音声対話システムを使用する状況が増加している．このとき，システムを初めて使うユーザは事前教示を受けておらず，システムに関する知識が十分ではない．以下では，このようなユーザを初心者ユーザと呼ぶ．音声対話システムが広く一般に用いられるようになるには，このような状況にも対処できる必要がある．初心者ユーザの発話は，システムが受理できない**システム想定外発話**を多く含み，音声認識誤りによるシステムの誤動作を引き起こす場合がある．システムがユーザの多様な発話をすべて言語理解できるよう，語彙や文法を網羅的に記述するのは事実上不可能であるため，システム想定外発話は不可避な問題である．

システム想定外発話は，マルチドメイン音声対話システムではさらに重要な課題となり，初心者ユーザがシステムを扱いにくい一因となっている．マルチドメイン音声対話システムは一般に，独立に設計された単一ドメインを統合して構築される．ここで**ドメイン**を，マルチドメインシステム内のサブシステムと定義する．このようなアーキテクチャでは，ユーザの要求がどのドメインでなされているかを推定する処理—**ドメイン選択**—が必要不可欠である．ドメイン選択誤りはユーザとの円滑な対話を著しく阻害するため，ドメイン選択には想定外発話に対する頑健性が求められる．従来のドメイン選択手法[1)–3)]は，発話の言語理解結果や対話履歴のみを利用しており，想定外発話からドメイン選択に有効な情報を取得できなかった．

本研究では，以下の 2 つのアプローチにより，想定外発話に対処する．

(1)　Web からの文書収集と Latent Semantic Mapping（LSM）[4)] を用いたトピック推定（3 章に対応）

(2)　トピック推定と対話履歴との統合（4 章に対応）

まず，『ユーザが本来意図していたドメイン』をトピックとして定義し，想定外発話に対してこれを推定する．これにより，想定外発話からもドメイン選択に有効な情報を取得でき

†1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

表 1　トピック推定結果と対話履歴の関係
Table 1　Relationship between topic estimation and dialogue history.

	想定外発話に対する頑健さ	文脈情報の考慮
トピック推定結果	○	×
対話履歴	×	○

る．次に，このトピック推定結果と，我々が以前に開発した手法で得られる対話履歴[1] を統合することで，想定外発話に頑健なドメイン選択器を構築する．我々は以前に，対話履歴をドメイン選択に反映させるために，現発話の音声認識結果に加えて，システムの内部状態やそれまでのユーザ発話に関する特徴を用いて，決定木学習によりドメイン選択を行った．**表 1** に示すように，トピック推定と対話履歴の利用は相補的な関係になっている．すなわち，トピック推定結果は想定外発話に対しても比較的信頼できるのに対し，対話履歴の利用は想定外発話に起因する言語理解誤りの悪影響を受ける．一方で，トピック推定は一発話から得られる情報のみを用いて行われるのに対して，対話履歴の利用は文脈を考慮できる．トピック推定と対話履歴の利用という相補的な手法を統合することで，より高精度なドメイン選択手法を開発する．

以下，2 章では，想定外発話に関する我々のシステム構成法であるアーキテクチャについて概観し，3 章では，Web からの学習データ収集法と LSM を用いたトピック推定について詳述し，4 章で，トピック推定と対話履歴の統合法について詳述する．5 章で，本システムの評価を行い，6 章で本論文をまとめる．

2.　マルチドメイン音声対話システムにおける想定外発話への対処

2.1　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は，バス運行案内やレストラン検索などの複数のタスクドメインを単一のシステムで扱える．このようなシステムは，ユーザの多様な要求に単一インタフェースで応答できるため，ユーザにとって利便性が高い．一方，多数の話題を扱う必要があるため，構築に多大な労力がかかるという問題がある．したがって，新たなドメインの構築の容易さや，構築したドメインをシステムに追加する容易さが求められる．本研究ではこれを**ドメイン拡張性**と呼ぶ．ドメイン拡張性を満たすために，マルチドメイン音声対話システム構築アーキテクチャとして個々のドメインを独立に統合してシステムを構築する手法が提案されている[2]．多くのマルチドメインシステム[2),5),6)] がこのアーキテクチャに基づいて開発されている．このアーキテクチャでは，システムは複数のドメインとそれらを統括するシステム本体からなる．システム本体は各ドメインの内部状態には関知しないため，ユーザの要求がどのドメインに対するものかを推定する処理—**ドメイン選択**—が必要不可欠となる．

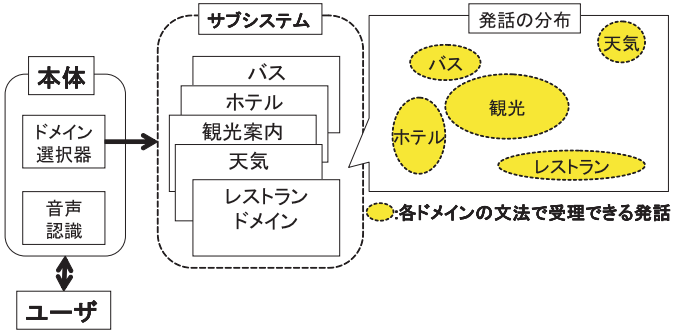

図 1　マルチドメイン音声対話システムのアーキテクチャ
Fig. 1　Architecture of multi-domain spoken dialogue system.

本研究が想定する 5 ドメイン音声対話システムのアーキテクチャ（**図 1**）を例にとり，具体的な処理を説明する．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，結果を各ドメインに送信する．各ドメイン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ドメインから送信された情報をもとに，応答すべきドメインの選択を行う．選択されたドメインは，対話状態に基づき次発話を決定する．システム本体はその情報を受け取って音声合成を行い，ユーザに出力する．

上述のアーキテクチャにおいて，ドメイン選択には以下の要求条件が存在する．本研究では，これを満たすドメイン選択器を構築する．

要求条件 1　拡張性を損なわないドメイン選択の枠組み
ドメイン選択も本アーキテクチャの一部となるため，ドメイン拡張性は同様に課せられる．すなわち，たとえばドメイン選択器の構築後に新たにドメインが追加された場合に，拡張後のシステムに少ない労力で選択器を適用できるフレームワークが要求される．

2.2　想定外発話への対処のためのトピック

マルチドメイン音声対話システムにおいて，初心者ユーザの発話は想定外発話になりやすいという問題がある．ここで，マルチドメインシステムのいずれのドメインにおいても受

理・解釈できない発話の集合を **“システム想定外発話”** と定義する．この問題は，マルチドメインシステムのアーキテクチャに起因する．まず，システムは独立に開発されたシステムを統合して構築されるため，各ドメインの言語理解文法が一貫しているとは限らない．このため，ユーザにとって各ドメインで受理される発話を推測しにくい．さらに，マルチドメインシステムでは扱うタスクが広く，ユーザの発話が多様になるからである．このため，ドメイン選択において，以下の要求条件を満たす必要がある．

要求条件 2　想定外発話に対する頑健性

想定外発話はドメイン選択誤りの原因となるため，ドメイン選択において重大な課題である．これは，想定外発話は言語理解誤りを引き起こすからである．音声対話システムでは，ある発話を受理・解釈できる言語理解文法が存在しない場合，そのような発話に対して正しい言語理解結果が得られない．想定外発話に対する頑健性として，具体的には以下の 2 点が要求される．

(1)　想定外発話に起因する誤ったドメイン遷移を防ぐ必要がある．正しいドメインが推定できない場合を検出し，発話を棄却すれば，システムの誤動作を防ぐことができる．

(2)　さらに，多くの発話に対して一意にドメインを決定できるのが望ましい．一意にドメインを決定できれば，具体的な応答が可能になる．

本研究では，想定外発話への対処として，ユーザの意図推定を行う．具体的には，システム想定外であっても，その発話内容に最も関連したドメインを **“トピック”** と定義し，これを推定する．ドメインとトピックの関係およびその具体例を**図 2** に示す．あるドメインの受理できる発話範囲は，システム開発者が用意した言語理解文法の範囲によって定められる．これに対して，あるトピックを指し示す発話の範囲は言語理解文法の範囲を越えて定義される．このため，トピックはつねにドメインを包含する．たとえば，「お手ごろ価格でおいしいランチがあるお店」という発話は，発話を理解するための文法規則が規定されておらずレストランドメイン外であっても，そのトピックはレストラントピックとなる．

要求条件 1 を満たすため，本研究ではドメインの拡張性を損なわないトピック推定を行う．具体的には，

(1)　学習データの Web からの大量収集

(2)　Latent Semantic Mapping（LSM）[4] を用いた学習データのノイズの影響の除去

というアプローチをとる．Web からの学習データの収集はあらゆるトピックに対して容易であり，収集に労力がかからないため，ドメイン拡張性を損なわない．その反面，求めるトピックに強く関連する文書のみを含むとは限らないため，事前に対話コーパスを用意した場合のような質の高い学習データは得られない．そこで，2 つ目のアプローチとして，LSM を用いる．LSM は，単語と文書との関係を低次元の空間に圧縮して写像することで，文書の潜在的な意味をベクトル空間として表現できる．これにより，学習データに混在するノイズの影響を取り除いたトピック推定が可能となる．

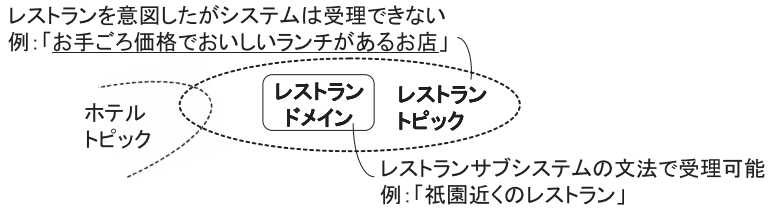

図 2　ドメインとトピックの関係およびその具体例
Fig. 2　Relationship between domains and topics.

トピック推定に関する関連研究として，対話コーパスからの学習により発話の話題を推定し，Support Vector Machine（SVM）や線形判別を用いることで，システムの扱っていない話題を検出する研究がある[7]．また，文献 8) では，トピック推定の方法として，SVM，Latent Semantic Analysis（LSA）およびトピック依存 N-gram の 3 種類の手法が比較評価されている．これらの手法には，あらかじめ収集された対話コーパスの存在を前提としており，ドメイン拡張性を満たさないという問題がある．初心者ユーザの想定外発話は多種多様であるため，頑健なトピック推定を行うためには，大量の学習データが必要となる．ドメインを追加する際に対話コーパスを新たに収集するのでは，多大な労力がかかるため，ドメイン追加が容易ではないからである．これに対して本研究では，Web から自動的に大量の学習データを収集することで，この問題に対処した．

想定外発話への対処に関する関連研究として，シングルドメインシステムを対象としたヘルプ生成の研究がある．Gorrel らは，想定外発話に対して，人手で分類した誤り原因を判別する決定木を対話データから学習している[9]．Hockey らは，2 つの音声認識結果を比べることで，想定外発話を発話区間誤り，未知語誤り，文法誤りの 3 つのクラスに判別している[10]．これらの研究は，想定外発話の誤り原因推定に基づきヘルプの内容を決定しており，Targeted Help と呼ばれる．これに対して本研究では，ユーザ発話のトピックを推定し，当該発話に対する応答生成をどのドメインに割り当てるべきかを判別している．マルチドメインシステムにおいて具体的なヘルプを提示するには，想定外発話の誤り原因を推定するだけ

では不十分であり，当該発話に応答すべきドメインを決定する必要がある．

2.3　システム想定外発話に頑健なドメイン選択

我々が開発したトピック推定と対話履歴は相補的な情報であり，これらの統合により，想定外発話に頑健なドメイン選択が可能となる．つまり，対話履歴は想定外発話による話題の遷移に追従できないのに対し，トピック推定は想定外発話からドメイン選択に有効な情報を取得できる．たとえば，「おいしいご飯が食べたいです」という想定外発話に対して，対話履歴は言語理解結果に基づいているため，言語理解誤りによる悪影響を受ける．このような発話に対して正しいドメインを選択するには，想定外発話に頑健なトピック推定を用いる必要がある．一方で，我々が開発したトピック推定は 1 発話のみに対して行われるため，文脈を考慮していない．たとえば，「お手ごろな値段のところがいいです」という発話は，本質的に文脈に依存する．そのため正しいドメインを選択するには，トピック推定だけでなく対話文脈を考慮する必要がある．

本研究では，システム想定外発話に対するトピック推定と対話履歴から得られる特徴量を用いてドメイン選択器を学習する．本手法の概略を**図 3** に示す．まず，ドメイン選択を以下の 4 クラス判別とした．すなわち，

(I)　1 つ前の応答を行ったドメイン
(II)　言語理解結果に関して最尤のドメイン
(III)　トピック推定用認識器の認識結果のトピック推定に対して最尤のドメイン
(IV)　それ以外のいずれかのドメイン

と定義した．これら 4 クラスは，ドメインが追加された際にも一様に定義でき，ドメイン拡張性を損なわない枠組みとなっている．次に，ドメイン選択器に用いる特徴量として，文献 1) で用いられていた，対話文脈に関する特徴量や言語理解結果に関する特徴量に加えて，トピック推定から得られる情報をともに用いる．これにより，想定外発話に対しても高精度なドメイン選択が可能となる．

磯部らは，トピックごとに音声認識器を用意し，その認識スコアを比較することでドメインを選択した[11]．言語モデルを，言語理解文法から生成された文書やトピック推定用に収集された文書から学習することで，選択肢 (II) や (III) を考慮したドメイン選択を実現できる．しかし，音声認識器のスコアを単に比較するだけでは，対話文脈を考慮することができない．

一方で，マルチドメイン音声対話システムにおいて対話履歴を反映させる場合には，しばしば直前の発話と同じドメインの継続が仮定される．Lin らはドメイン選択に際して前ターンのドメインを選択しやすくする制約を設けた[2]．O'Neill らはタスクが達成されるまでドメイン遷移を許さない戦略をとった[3]．Lane らは想定外発話検出の特徴量として，前発話までの特徴量や検出結果をともに用いて判別を行った[12]．これらは，ユーザが同じ話題での発話を続ける限りは有効な戦略であるが，対話中のトピックの変化や前発話までのドメイン選択誤りを想定していない．これに対して我々は以前，トピックの変化や直前のドメイン選択の信頼度を表す特徴量を導入して，ドメイン選択に対話履歴を反映させた[1]．しかし依然，想定外発話からドメイン選択に有効な情報を得ていないため，想定外発話に対する音声認識誤りの影響を受けやすいという問題点があった．さらには，そのような音声認識誤りを正しく棄却した場合でも，システム応答を行うドメインを一意に選択できない．

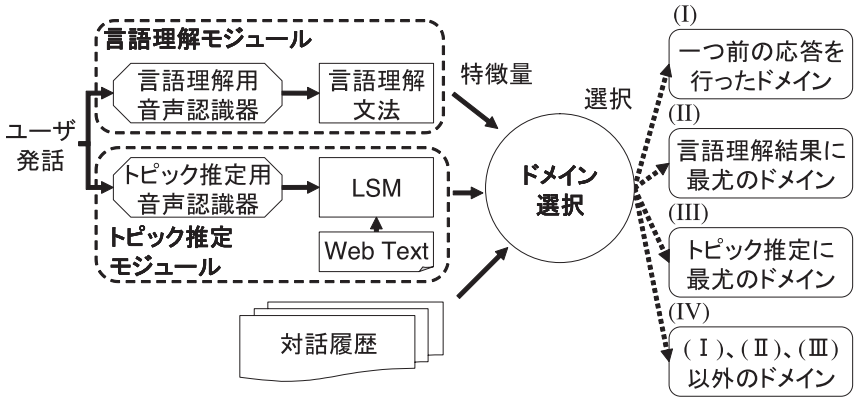

図 3　ドメイン選択の概略
Fig. 3　Overview of our domain selection.

本手法では，従来のドメイン選択手法[1)–3)] と比べて以下の 2 点の改善が期待される．

(1)　ドメイン選択精度の改善
本手法では，ドメイン選択にトピック推定結果を考慮している．想定外発話から得られる有効な情報をドメインの判別に導入することで，想定外発話に対してもドメイン選択精度の向上が見込める．

(2)　具体的な正解ドメインを選択できる発話の増加
本手法では，選択肢 (III) を導入することで，より多くの発話に対して一意に応答すべきドメインを決定できる．手法 1) では，想定外発話は (IV) もしくは誤ったクラスに判別される．これにより，想定外発話に対して具体的な応答が可能になる．

U1：**明日の京都の天気を教えて** (ドメイン：天気)
S1：明日の京都の天気は，晴れです．
U2：**京都の 夜景がきれいな場所** (ドメイン：観光)
(下線部が文法外．「京都外大前より競馬場」と誤認識)
S2 (言語理解結果のみを利用)：京都外大前から競馬場までですか？ (ドメイン：バス)
S2 (文献 1) の手法)：理解できませんでした． (ドメイン：その他)
S2 (本手法)：理解できませんでした．観光については，場所，観光施設タイプなどが指定できます．たとえば，「祇園周辺の寺を検索」などとおっしゃってください．(ドメイン：観光)

図 4　システム想定外発話を含む対話例
Fig. 4　Example of dialogue including out-of-grammar utterances.

本手法により可能となる対話例を**図 4** に示す．U2 でユーザは観光に関する発話を行うが，想定外発話であったため正しく認識されず，言語理解結果に対する最尤のドメインはバスドメインとなる．このとき，正解である観光ドメインは，1 つ前に応答を行ったドメイン（天気ドメイン）でも言語理解結果に対して最尤のドメイン（バスドメイン）でもない．S2（文献 1) の手法）では，「その他のドメイン」という正しい判別結果を得たとしても，具体的なドメインが得られず，具体的な応答ができない．これに対して本手法では，S2（本手法）のように，トピック推定を用いることで，システムの言語理解部では解釈できない表現を含む発話に対しても正解ドメインを推定できる．想定外発話に対しては言語理解結果が信頼できないため，S2（本手法）では言語理解結果を棄却し，当該ドメインに応じたヘルプを提示している．このように，対話の履歴から得られる情報に加えて，トピック推定から得られる情報を統合することにより，ドメイン選択精度の向上とともに，ドメインに応じた具体的な応答を行えるようになる．

3. Web からの大量文書の自動収集と LSM に基づくトピック推定

トピック推定は，ユーザ発話と Web 収集の学習データとの近さを LSM を用いて計算することで行う．トピック推定の概略を図 3 のトピック推定モジュールに示す．以降，レストラン，観光，バス，ホテル，天気の 5 ドメインを扱うマルチドメインシステムを例として説明を進める．このシステムには，それぞれのドメインに対応する 5 つのトピック（レストラン，観光，バス，ホテル，天気）と，「はい」や「もどる」など，どのドメインにも共通する発話の集合に対応するコマンドトピックがある．トピック推定は，以下の 2 つの手順により行われる．

表 2　ホテルトピックのキーワード
Table 2　Keyword sets for the hotel topic.

ホテル 泊ま
旅館 泊
旅館 一泊
ホテル 泊
ホテル 部屋

3.1 学習データの収集

コマンド以外の 5 つのトピックに関して，各トピックにつき 10 万文をツール[13] を用いて Web から収集した．このツールにより，具体的に以下の処理が行われる．まず，人手で 10 個前後の検索キーワードを指定し，5,000 程度の Web ページから文書を収集する．たとえばホテルに関するキーワードは，**表 2** に示した 6 つのキーワードセットを用いた．これらのキーワードは，各トピックごとの検索結果が互いに大きく異なるように選定した．また，システムの機能を考慮し，その機能を扱う言語表現を含む検索結果が収集できるようにした．たとえば，ホテルトピックのキーワードに「予約」が入っていないのは，本システムのホテルに関するタスクが検索のみであり，予約タスクを扱っていないからである．

次に，収集した Web 文書から，学習データとして適したものを選択する．文書選択のために，Wikipedia[⋆1]から文書を集め，数百文程度の文書集合（seed data）を作成する．seed data は，トピックに関連のある Wikipedia のページの文章をそのままコピーすることで構築した．ホテルトピックにおいては，seed data の単語総数は 9,517 であった．次に，seed data をもとに統計的言語モデルを作成し，その言語モデルをもとに収集した Web ページの文書に対してパープレキシティを計算する．統計的言語モデルには，trigram モデルを用いた．取得した Web ページの文書のすべてが学習データとして適しているとは限らないため，単語パープレキシティを用いてフィルタリングを行う．Web から収集したテキストの各文に対して，このモデルによる尤度を計算する．その尺度として，単語パープレキシティを用いた．単語パープレキシティは，ある単語 1 個が出現する確率の相乗平均の逆数で定義される．パープレキシティの低いものから順に 10 万文を学習データとして選択する．ホ

⋆1 http://ja.wikipedia.org/

テルトピックにおいては，単語パープレキシティが 1,156.83 以下となる文のみを学習データとして選択した．コマンド発話の学習データに関しては Web から収集するのは困難であるので，175 文を人手で準備した．また，システムの言語理解用文法から各トピックにつき 1 万文を生成し，学習文書に加えた．この際，システムの言語理解用文法から生成した文の集合と，Web から収集した文の集合のバランスを考慮し，前者の頻度を 3 倍した後に混合し，学習文書とした．

以上の作業で各ドメインごとに収集した 13 万文をランダムに d 個に分割し，学習文書を構成した．各トピックを 1 点で表現するのではなく，複数の点として表現することでトピックの広がりを表現した．これらの学習文書は対話コーパスを収集するほどの多大な労力を必要とせずに収集可能であるため，ドメイン拡張性を損なわないという利点がある．一方で，求めるトピックに強く関連する文書のみを含むとは限らず，ノイズを含んでいるという問題点もある．

3.2　LSM を用いたトピック推定

各トピックに対する学習文書集合と入力発話との近さを計算することで，トピック推定を行う．本研究では，学習データに含まれるノイズの影響を除去するために，LSM[4] を適用する．LSM は自然言語処理分野において，文書分類や要約[14] などに用いられる．LSM を用いたトピック推定手法を以下に示す．

まず，各学習文書に対する単語の頻度をもとに得られる $M \times N$ 共起行列 W を求める．ここで，M は学習文書集合に現れる異なり単語数，N は学習文書数である．また，推定の対象とするトピック数を n，トピックごとの学習文書数を d とすると，$N = n \times d$ と表される．ここで，d は 3.1 節で学習文書集合の分割に用いたものと同一である．具体的には，共起行列 W の (i, j) 成分 $w_{i,j}$ の求め方は，以下の式で与えられる．

$$w_{i,j} = (1 - \varepsilon_i)\frac{\kappa_{i,j}}{\lambda_j} \tag{1}$$

ここで，$\kappa_{i,j}$ は学習文書 doc_j に現れる単語 r_i の出現回数，λ_j は学習文書 doc_j の単語数である．また，ε_i は学習文書全体における単語 r_i のエントロピーである．

その共起行列に対して特異値分解と次元縮約を行い，共起行列の階数を k に減じる．この次元縮約により，学習データのノイズの影響を除去する．まず，W に特異値分解を行い，行列 U，S，V を得る．

$$W = USV^T \tag{2}$$

ここで，W の階数を $rank$ とすると S は特異値 $s_1 > s_2 \cdots > s_{rank}$ からなる対角行列である．次に，$s_1, s_2, \cdots, s_k$ のみを使い，W を $\hat{W}$ で近似することで，次元縮約を行う．ただし，$k < rank$ とする．

$$W \approx \hat{W} = \hat{U}\hat{S}\hat{V}^T = \begin{bmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{M1} & \cdots & u_{Mk} \end{bmatrix} \begin{bmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_k \end{bmatrix} \begin{bmatrix} v_{11} & \cdots & v_{N1} \\ \vdots & \ddots & \vdots \\ v_{1k} & \cdots & v_{Nk} \end{bmatrix} \tag{3}$$

以上で得られたを $\hat{W}$ もとに，N 個の学習文書それぞれに対して k 次元空間のベクトル表現を得る．具体的には，学習文書 doc_j の k 次元空間でのベクトル表現 $\bar{v}_j$ は，

$$\bar{v}_j = \hat{v}_j\hat{S} \tag{4}$$

によって与えられる．ここで，$\hat{V}$ の列ベクトルを $\hat{v}_j$（$1 \leq j \leq N$）とする．本研究で作成した共起行列は，$M = 67{,}533$，$N = 120$，$n = 6$，$d = 20$ である．次元縮約に関しては $k = 50$ とした．

次に，入力発話の認識結果に対しても，k 次元ベクトル $\bar{v}_{input}$ を求める．入力発話の音声認識には，システムの言語理解に用いる言語モデルとは別に，3.1 節で収集したデータをもとに作成した統計的言語モデルを用いる．本研究では，実時間で $\bar{v}_{input}$ を計算するために，近似手法を採用する[4]．まず，音声認識結果に対応する M 次元ベクトル c を，式 (1) を用いて求める．次に，M 次元ベクトル $v_{input} = c^T U$ を計算する．$\bar{v}_{input}$ は，v_{input} の k 次元成分までを用いることで得られる．

トピックに属する d 個の学習文書と入力発話とのコサイン距離の最大値を，トピックと入力発話の近さと定義する．この尺度に基づき，入力発話に最も近いトピックを，推定結果として出力する．

4.　トピック推定と対話履歴の統合によるドメイン選択

本研究では，トピック推定結果と対話履歴の統合により，想定外発話に頑健でドメイン拡張性を損なわないドメイン選択を行う．図 3 に示すように，ユーザ発話の言語理解結果や対話履歴，トピック推定から得られる特徴量を入力として，ドメインを判別する決定木を対話データから学習する．本手法は，ドメイン選択の特徴量にトピック推定結果を用いており，想定外発話に対して頑健である．また，ドメイン拡張性を損なわないために，特定のドメイ

ンに依存しない特徴量を用い，出力とするラベルも個々のドメインに依存しないように設計している．

まず，決定木が判別するクラスについて述べる．本研究では，対話データに対して，正解クラスを発話ごとにラベル付けし，決定木を学習する．図 3 に示すように，以下の 4 クラスを定義した．この際，正解クラスが一意に定義できるように設計した．

クラス (I)： 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

クラス (Ⅱ)： (I) 以外の場合で，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

クラス (Ⅲ)： (I)，(Ⅱ) 以外の場合で，正解ドメインが，トピック推定結果に対して最尤のドメインである場合

クラス (Ⅳ)： その他の場合

これらの選択肢は，個々のドメインの判別ではなく，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，得られた選択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．まず，対話履歴，言語理解結果，トピック推定結果に対応して，(I)，(Ⅱ)，(Ⅲ) を定義した．また，(Ⅳ) は，ドメイン選択には誤りが不可避である点を考慮して定義した[1)]．これにより，1 つ前のユーザ発話でドメイン選択誤りが起こった状況を検出でき，誤ったドメインを選択し続ける問題を回避できる．たとえば，(Ⅳ) が選択された場合は，1 つ前で応答したドメインへの遷移を禁止したうえで，1 つ前のユーザ発話のドメイン選択をやりなおすといった対話戦略が考えられる．

次に，ドメイン選択のために利用する特徴量について述べる．本研究では，4 種類の特徴量を設計する．1 つ前の応答を行ったドメインに関する特徴量（**表 3**），言語理解用音声認識器による認識結果に関する特徴量（**表 4**），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（**表 5**）は，文献 1) で用いられたものである．これらに加えて，トピック推定に関する特徴量（**表 6**）を新たに導入する．これにより，システム想定外発話に対しても正しいトピックが推定可能となる．

以下，新たに導入した表 6 の特徴量について詳しく述べる．T1〜T6 は，トピック推定結果がどの程度信頼できるかの指標である．ここで，トピック T の信頼度は，$CM_T = closeness_T / \sum_t closeness_t$ として定義する．t はシステムに存在するドメインであり，$closeness_t$ はトピック t と入力発話の近さである．T1 と T3 は，当該トピックに認識結果がどれだけ関連しているかを表現するために定義した．これに対して，T2 と T4 は，当該トピックに対してどれだけ文脈に依存した発話であるかを表現するために定義した．たとえば，「上限予算 5 千円」という発話は，レストランとホテルの両方に強い関連があり，それぞれのトピックに対する近さは高い値を，信頼度は低い値を示す．このように，近さが高い値を示すのにもかかわらず，トピック信頼度が低い発話は文脈に強く依存していると判断できる．さらに，T5 と T6 のように，近さと信頼度の差を定義することで，(I) と (Ⅲ) のどちらを優先すべきかを判断している．次に，ラベル (I)，(Ⅱ)，(Ⅲ) の関係を表すために，T7〜T9 を導入した．たとえば，トピック推定で最尤のドメインと 1 つ前のドメインが一致する場合は，1 つ前のドメインはより信頼できると考えられるからである．T10 は，音声認識結果があまりに短い発話のトピック推定結果は信頼できない場合が多いという傾向があるため定義した．また，T11〜T13 を定義することで，ユーザ発話が想定外発話かどうかの情

表 3　1 つ前の応答を行ったドメインに関する特徴量
Table 3　Features representing confidence in the previous domain.

P1：	そのドメインに遷移した後のユーザの肯定応答回数
P2：	そのドメインに遷移した後のユーザの否定応答回数
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：	スロットの変化の度合（=P5/P6）
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（=P2/(P1+P2)）
P9：	対話におけるユーザの否定応答の割合（=P2/P6）
P10：	タスクの状態

表 4　言語理解用音声認識器による認識結果に関する特徴量
Table 4　Features of ASR results.

U1：	(I) で言語理解できた音声認識結果の音響スコア
U2：	(I) で言語理解できた音声認識結果の文としての事後確率
U3：	(I) で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
U4：	(Ⅱ) が受理した音声認識結果の音響スコア
U5：	(Ⅱ) が受理した音声認識結果の文としての事後確率
U6：	(Ⅱ) 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
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）

表 5　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
Table 5　Features representing situations after domain selection.

C1：	(I) を選択した場合の，そのドメインのタスクの状態
C2：	(I) で言語理解した場合，肯定応答かどうか
C3：	(I) で言語理解した場合，否定応答かどうか
C4：	(I) で言語理解した場合，変化するスロット数
C5：	(I) で言語理解した場合，変化する共有スロット数
C6：	(II) を選択した場合の，そのドメインのタスクの状態
C7：	(II) で言語理解した結果が，肯定応答かどうか
C8：	(II) で言語理解した結果が，否定応答かどうか
C9：	(II) を選択した場合に変化するスロット数
C10：	(II) を選択した場合に変化する共有スロット（地名）数
C11：	(II) が，それまでに存在したか

表 6　トピック推定に関する特徴量
Table 6　Features of topic estimation result.

T1：	(III) に対応するトピックと発話の認識結果との近さ
T2：	(III) に対応するトピックの信頼度
T3：	(I) に対応するトピックと発話の認識結果との近さ
T4：	(I) に対応するトピックの信頼度
T5：	ユーザ発話と (I)，(III) の近さの差（=T1−T3）
T6：	(I) と (III) のトピック信頼度の差（=T2−T4）
T7：	(III) と (II) が一致するか
T8：	(III) と (I) が一致するか
T9：	(III) がコマンドトピックかどうか
T10：	トピック推定用音声認識器による認識結果の長さ（音素数）
T11：	トピック推定用音声認識器による認識結果の音響スコア
T12：	T11 と U1 の一音素あたりの音響尤度差（=(T11−U1)/T10）
T13：	T11 と U4 の一音素あたりの音響尤度差（=(T11−U4)/T10）

報を表す[15]．ユーザ発話が想定外発話であれば，ラベル (II) よりも (III) の方が信頼性が高いと考えられる．

5. 評価実験

本論文では，トピック推定のみの推定精度についてまず 5.2 節で評価する．続いて 5.3 節でそれを用いたドメイン選択精度の評価を行う．

5.1 評価対象の対話データ

評価データには，2 種類の対話データを用いた．以降それぞれを**教示あり対話データ**，**教示なし対話データ**とする．

教示あり対話データは，被験者 10 名から収集した 2,191 発話からなり，文献 1) でのドメイン選択精度の評価に用いられたデータである．被験者は，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．データ収集時のシステムは，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，1 つ前の応答を行ったドメインには，音響尤度に 40 を加算して比較している[1]．被験者は練習を行った後に文献 1) のシステムを用いているので，システム想定内発話が多く含まれる．

一方で，教示なし対話データは，模擬対話により収集された，被験者 8 名の 272 発話である．システムが受理できる発話パターンなどの教示を与えないでデータを収集したため，この対話データを教示なし対話データと呼ぶ．そのため，事前教示を受けていない一般ユーザによるシステムの使用条件に近く，想定外発話が多い．

音声認識は，言語理解用とトピック推定用の 2 つの認識器を用いて行った．言語理解用音声認識器には Julian[16] 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と同等であり，語彙サイズは 7,373 である．トピック推定用音声認識には Julius[16] を用いた．言語モデルは，トピック推定の際に使用した学習データを用いて構築した trigram モデルである．また，学習に用いた文書の単語総数は 11,866,432 であった．語彙サイズは 56,453 である．音響モデルは 3,000 状態不特定話者 PTM トライフォンモデル[16] を用いた．教示あり対話データにおける単語正解率は，言語理解用音声認識器（Julian）で 63.3%，トピック推定用音声認識器（Julius）で 69.6%であった．教示なし対話データにおいては，それぞれ 26.6%，67.3%であった．なお，トピック推定用音声認識器の方が単語認識率は高いが，その認識結果がシステムの言語理解部で受理できるキーワードやフレーズであるとは限らないため，現状では Julian の認識結果を言語理解に用いている．Julius の音声認識結果に対する頑健な言語理解は，今後の課題である．

また，教示なし対話データはトピック推定の評価のみに用いる．このデータは模擬対話により収集されたため，初心者ユーザによるシステムの使用条件に近く，想定外発話が多く含まれる．このため，教示なし対話データを用いることで，想定外発話に対するトピック推定の頑健性を評価できる．一方，教示なし対話データはシステムを用いずに収集されており，システムの内部状態・履歴に関する特徴量が取得できない．このため，ドメイン選択の評価には用いていない．

5.2　トピック推定精度の評価

上述の 2 つの対話データに対して，トピック推定精度を評価した．実験条件を以下に述べる．2 種類の対話データに対して，人手でトピック推定の正解ラベルを与えた．正解ラベルとして．レストラン，ホテル，観光案内，バス運行情報案内，天気情報，コマンド，トピックが不明な発話の 7 つのいずれかを付与した．この際，発話断片や雑音のみからなる 63 発話を取り除いた，計 2,400 発話に対してラベル付けを行った．トピックが不明な発話という正解ラベルを用意したのは，1 発話のみからではトピックを決定できない場合があるからである．たとえば，「上限予算五千円」という発話は，その発話のみではホテルトピックとレストラントピックの両方の可能性があり，文脈によってトピックが異なる．また，ホテルやレストランの検索システムに対する「昨日のプロ野球の結果を教えて」という発話は，システムが持つどのトピックにも該当しない．本論文では，トピック信頼度がある閾値 θ 以下の場合に，トピックが不明な発話であると判定した．

学習データの Web 収集，LSM による次元縮約ともに用いない場合をベースライン手法として，トピック推定の精度を比較評価した．

ベースライン手法 1：各ドメイン文法により生成された文書のみを学習データとし，ベクトル空間モデル[17] を用いてトピックを推定する．具体的には，語彙サイズ次元からなるベクトル空間を構成し，出現単語の頻度を要素として持つベクトルで音声認識結果やトピックを表現する．トピック推定結果は，音声認識結果のベクトルに対して最もコサイン距離の小さいトピックとして出力する．トピックを表すベクトルは，ドメイン文法から生成された文の集合内の出現頻度をもとに計算する．このベースラインは，各トピックに近い単語の数が最も多いトピックを出力とするような，単純なルールベース手法におおよそ対応する．

ベースライン手法 1，Web 収集のみを用いた手法，LSM のみを用いた手法，本手法の 4 つの場合でトピック推定正解率を比較評価した．ここで，ベースライン手法 1 の場合の言語モデル作成には，Web 収集データを使用している．これは，すべての手法で同一の音声認識結果を用い，純粋にトピック推定手法の違いによる推定精度の変化を比較評価するためである．トピック推定の正解率を**表 7** に示す．また，すべての発話に対して本手法を適用した場合の Confusion matrix を**表 8** に示す．ここで，θ はそれぞれの場合で，10-fold cross validation により決定した．

ベースライン手法 1 では，教示なし対話データに対する正解率が教示あり対話データより 19.8 ポイント低い．これは，教示なし対話データには想定外発話が多く含まれるからである．Web 収集のみを用いた手法では，Web から大量の学習データを収集し，システムの知識を拡張している．このため，教示なし対話データに対して正解率が 7.4 ポイント改善された．一方で，Web データにはノイズが含まれるため，教示あり対話データに対しては 7.3 ポイント悪化している．これに対して，LSM と Web 収集の両方を導入する本手法では，2 つの対話データそれぞれにおいて最も良い精度でトピックを推定している．この大幅な正解率の改善は，Web から収集したデータに含まれるノイズの影響を，LSM を用いて取り除いているためである．これらの事実は，

- 学習データの Web 収集と LSM の両方を用いることによる相乗効果
- 想定外発話に頑健なトピック推定の実現

を示している．

表 7　各手法におけるトピック推定精度
Table 7　Correctness of topic estimation for each method.

	教示あり対話データ	教示なし対話データ
ベースライン手法 1	56.2%	36.4%
＋ (1) Web 収集	48.9%	43.8%
＋ (2) LSM	61.3%	44.9%
＋ (1) ＋ (2)（本手法）	62.2%	59.6%

表 8　本手法によるトピック推定の Confusion matrix
Table 8　Confusion matrix in topic estimation based on our method.

	正解ラベル								
出力	レストラン	ホテル	観光	バス	天気	コマンド	不明	計	（精度）
レストラン	128	8	10	3	8	1	114	272	(0.47)
ホテル	3	144	5	2	5	0	28	187	(0.77)
観光	5	14	138	4	7	1	90	259	(0.53)
バス	4	1	19	65	4	2	43	138	(0.47)
天気	5	1	3	3	140	1	22	175	(0.80)
コマンド	7	13	35	1	22	574	183	835	(0.69)
不明	27	47	98	20	43	3	296	534	(0.55)
計	179	228	308	98	229	582	776	2,400	(0.619)

5.3　ドメイン選択の評価

教示あり対話データに対する本手法のドメイン選択精度を比較評価した．

以下に実験条件を述べる．決定木の構築には C5.0[18] を用いた．特徴量は，Backward stepwise selection により選択したものを用いる．本手法に関しては，表 3，4，5，6 に示

表 9　特徴選択の結果得られた特徴量
Table 9　Surviving features after feature selection.

本手法	P2, P3, P4, P5, P6, P7, P9, P10, U2, U3, U5, U6, C3, C6, C8, C10, C11, T2, T3, T4, T5, T7, T8, T9, T10, T11, T12
ベースライン手法 2	P1, P4, P5, P8, P9, P10, U1, U2, U3, U5, U6, U7, U9, C8, C9, C11

表 10　本手法でのドメイン選択の Confusion Matrix
Table 10　Confusion matrix in four-class classification.

		識別結果				
		(I)	(II)	(III)	(IV)	計（再現率）
正解	1 つ前 (I)	1,348	34	23	37	1,442 (93.5%)
	言語理解最尤 (II)	93	258+10†	14	5	380 (67.9%)
	トピック推定最尤 (III)	81	7	37	6	131 (28.2%)
	その他 (IV)	130	11	13	84	238 (35.5%)
計（適合率）		1,652 (81.6%)	310 (83.2%)	87 (42.5%)	132 (63.6%)	2,191 (78.8%)

†：複数のドメインで最も高いスコアが得られた場合のランダムな選択による 10 の誤りを含む

した 43 の特徴量から選択を行った．選択された特徴量を**表 9** 上段に示す．評価は 10-fold cross validation を用いて，発話ごとに行った．また，最高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．教示あり対話データは，2.3 節で述べた (I)，(II)，(III)，(IV) のいずれかを発話ごとに人手でラベル付けされている．

5.3.1　ドメイン選択精度の評価

本手法におけるドメイン選択精度を示す．**表 10** は，本手法における正解ラベルと識別結果の Confusion Matrix である．対角成分が正しく判別された数を表す．本手法におけるドメイン選択誤り数は 464（= 2,191 − 1,348 − 258 − 37 − 84）となった．トピック推定結果の導入による影響をより詳しく見ると，正解ラベルが (III) の発話のうち，37 発話が正しく選択されているが，この 37 発話は従来手法では正しいドメインを選択することができない発話である．たとえば，この 37 発話には，「京大正門前へ行くバス は動いていますか」（下線部が文法外）などの，システム想定外発話が含まれていた．また，ラベル (III) の再現率が 28.2%と低いのは，ラベル (III) の総発話数がラベル (I) に比べて大幅に少ないため，ほとんどの発話を (I) に識別するよう決定木が学習されたことが要因と考えられる．

また，ドメイン選択の際に有効だった特徴量を調査した．ここでは，特徴量を 1 つ取り除いた後にドメイン選択誤り数がどれだけ増加するかを調査した．ドメイン選択誤りの増加が上位 10 個の特徴量とその増加数を**表 11** に示す．トピック推定に関する特徴量が上位 10 個のうち 4 個（T7，T2，T10，T12）を占めており，トピック推定から得られる情報が効果的であることを示している．

表 11　各特徴量を除いた場合のドメイン選択誤りの増加数
Table 11　Increase in number of errors when feature was removed.

特徴量	U8	P9	T7	U6	T2	C8	U3	P5	T10	T12
誤り増加数	86	67	62	58	47	43	40	40	37	33

表 12　3 クラス判別における Confusion Matrix（ベースライン手法 2/本手法）
Table 12　Confusion matrix in three-class classification (Baseline/Our method)

		識別結果			
		(I)	(II)	(III)+(IV)	計（再現率）
正解	1 つ前 (I)	1,303/1,348	71/34	68/60	1,442 (0.90/0.93)
	言語理解最尤 (II)	104/93	238+14†/258+10†	24/19	380 (0.63/0.68)
	その他 (III)+(IV)	191/211	47/18	131/140	369 (0.36/0.38)
計（適合率）		1,598/1,652 (0.82/0.82)	370/320 (0.64/0.81)	223/219 (0.59/0.64)	2,191 (0.764/0.797)

†：複数のドメインで最も高いスコアが得られた場合のランダムな選択による 10 の誤りを含む

5.3.2　ベースライン手法 2 との比較

トピック推定を用いない場合をベースライン手法として，ドメイン選択精度を比較評価した．

ベースライン手法 2：文献 1) での提案手法によりドメイン選択を行う．すなわち，トピック推定に関する特徴量や出力を取り除いた場合に相当する．対話データのラベル (III) と (IV) を同一（その他のドメイン）と見なした後，3 クラス判別の決定木を学習した．ドメイン選択器の構築に用いた特徴量は，表 6 にあるものを用いず，表 3，4，5 に示した特徴量から選択した．選択された特徴量を表 9 下段に示す．

表 12 は，ベースライン手法 2 における正解尺度（3 クラス判別）を適用した場合の，本手法とベースライン手法 2 のドメイン判別結果を示している．表 10 と同様に対角成分が正しく判別された数を表す．表中の各セルの左側の数字はベースライン手法 2 の出力結果，右側は本手法による出力結果を表す．本手法による出力は表 10 と同じであるが，(III) と (IV) のラベルについては表 10 の対応するセルの和になっている．また，**表 13** は，本手法とベースライン手法のドメイン判別精度の比較結果を示している．ドメイン選択誤り数，具体的なド

表 13　本手法とベースライン手法 2 との比較
Table 13　Comparison between our method and baseline method.

	3 クラス判別における ドメイン選択誤り数 （誤り率）	具体的なドメインが 正しく選択できた発話数 （再現率）
本手法	445 (20.3%)	1,643 (75.0%)
ベースライン手法 2	519 (23.7%)	1,541 (70.3%)

メインが正しく選択できた発話数という 2 つの観点から，ドメイン選択精度を比較評価した．表 10 より，ベースライン手法 2 のドメイン選択誤り数は 519（= 2,191 − 1,303 − 238 − 131）であり，本手法では 445（= 2,191 − 1,348 − 258 − 140）である．また，具体的なドメインが正しく選択できた発話数は，「その他のドメイン」以外のラベルにおける正解発話数を求めることで算出した．本手法においては，表 10 より 1,643 発話（= 1,348 + 258 + 37）であり，ベースライン手法 2 においては，表 12 より 1,541 発話（= 1,303 + 238）であった．

まず，ベースライン手法 2 における正解尺度を適用した場合の，ドメイン選択誤り数の比較結果について述べる．表 13 より，ベースライン手法 2 におけるドメイン選択誤り数は 519 で，ドメイン選択誤り率は 23.7%（= 519/2,191）である．一方，ベースライン手法 2 における正解基準を適用した場合における本手法のドメイン選択誤り数は，表 13 より 445 となり，ドメイン選択誤り率は 20.3%（= 445/2,191）である．ドメイン選択誤り削減率は 14.3%（= 74/519）となる．表 12 によると，すべてのクラスにおいて，正解率の改善が見られる．(I) や (II) の場合も正解率の改善が見られるのは，T7 や T8 の特徴量が，(I) と (II) の判別において効果的な情報となったためと考えられる．

次に，システムが一意に正しいドメインを決定できた発話数を比較する．ベースライン手法 2 では，1 つ前の応答を行ったドメイン（ラベル (I)）と言語理解結果に対応するドメイン（ラベル (II)）で，一意に応答するドメインが選択できる．これに対して本手法では，ラベル (III) に相当する発話でも，トピック推定により応答すべきドメインが選択できる．表 13 によると，具体的なドメインが正しく選択できた発話数は，ベースライン手法 2 において 1,541 発話であり，再現率は 70.3%（= 1,541/2,191）である．一方で，本手法では表 10 より 1,643 発話であり，その再現率は 75.0%（= 1,643/2,191）である．よって，システムが一意に正しいドメインを決定できない発話の削減率は 15.7%（= 102/650）である．これは，ベースライン手法 2 より本手法の方が，より広範囲のユーザ発話に対して具体的なドメインを推定できることを示している．

6. 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，システム想定外発話に対しても，応答すべきドメインを頑健に選択する手法について述べた．10 名の被験者から収集した対話データ[1] を用いた評価実験により，従来手法と比べドメイン選択誤りが 14.3%削減されることを確認した．

本研究の意義を以下に述べる．

(1) マルチドメイン音声対話システムにおける想定外発話に頑健なドメイン選択の重要性を指摘し，これを実現した．これまでは，対話履歴と発話の言語理解結果のみを考慮しており，想定外発話からはドメイン選択に有効な情報を取得できなかった．本研究は，トピック推定により想定外発話に対しても有効な情報を取得し，さらに相補的な情報である対話履歴と統合した．これにより，想定外発話に頑健なドメイン選択手法を開発した．さらに，想定外発話に対しても正しいドメインを一意に決定することで，より多くの発話に対して具体的な応答を可能にした．

(2) 想定外発話に頑健なドメイン選択手法を，ドメイン拡張性を損わない枠組みで実現した．マルチドメインシステムは構築に多大な労力がかかるため，ドメインの追加や修正が容易に可能であることが求められる．ドメイン選択に関しても同様に，その再利用性が求められる．そのため本研究では，トピック推定を行う際に Web 収集データを利用した．これにより，学習データを人手で収集する必要がなく，ドメイン拡張性を損なわないトピック推定が可能となった．次に，トピック推定と対話履歴の利用を，ドメイン拡張性を損わない枠組みで統合した．具体的には，特定のドメインに依存しない特徴量を用い，特定のドメインを出力ラベルとしないドメイン選択器を構築した．これにより，ドメイン選択器の構築後に新たにドメインが追加された場合でも，少ない労力で拡張後のシステムに適用可能である．

現在，本論文において開発したドメイン選択手法が，ドメイン数を増加させた場合でも有効であることを実験的に検証中である．また，今後は本手法を実際のシステムへと実装し，その有効性を評価する．

謝辞　LSM の学習データの収集には京都大学河原研究室で開発された Webcollect[13] を用いた．また，評価用対話データは，ホンダ・リサーチ・インスティチュート・ジャパンの中野幹生氏らとの共同研究において，神田直之氏らとともに構築したシステムにより収集した．本研究の一部は，科研費，グローバル COE，SCAT 研究助成の援助を受けた．

参 考 文 献

1) 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：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，情報処理学会論文誌，Vol.48, No.5, pp.1980–1989 (2007).
2) Lin, B., Wang, H. and Lee, L.: A Distributed Agent Architecture for Intelligent Multi-Domain Spoken Dialogue Systems, *Proc. ASRU* (1999).
3) O'Neill, I., Hanna, P., Liu, X. and McTear, M.: Cross Domain Dialogue Modelling: An Object-Based Approach, *Proc. ICSLP*, Vol.I (2004).
4) Bellegarda, J.R.: Latent Semantic Mapping., *IEEE Signal Processing Mag.*, Vol.22, No.5, pp.70–80 (2005).
5) Markku, T. and Jaakko, H.: Jaspis2 – An Architecture for Supporting Distributed Spoken Dialogues, *Proc. Eurospeech*, pp.1913–1916 (2003).
6) Pakucs, B.: Towards Dynamic Multi-Domain Dialogue Processing, *Proc. Eurospeech*, pp.741–744 (2003).
7) 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).
8) Lane, I., Kawahara, T., Matsui, T. and Nakamura, S.: Out-of-domain detection based on confidence measures from multiple topic classification, *Proc. ICASSP*, Vol.1, pp.757–760 (2004).
9) Gorrell, G., Lewin, I. and Rayner, M.: Adding Intelligent Help to Mixed-Initiative Spoken Dialogue Systems, *Proc. ICSLP*, pp.2065–2068 (2002).
10) Hockey, B.A., Lemon, O., Campana, E., Hiatt, L., Aist, G., Hieronymus, J., Gruenstein, A. and Dowding, J.: Targeted Help for Spoken Dialogue Systems: intelligent feedback improves naive users' performance, *Proc. EACL*, pp.147–154 (2003).
11) 磯部俊洋，伊藤克宜，武田一哉：複数の認識器を選択的に用いる音声認識システムのためのスコア補正法，電子情報通信学会論文誌，Vol.J90-D, No.7, pp.1773–1780 (2007).
12) Lane, I. and Kawahara, T.: Incorporating dialogue context and topic clustering in out-of-domain detection, *Proc. ICASSP*, Vol.1, pp.1045–1048 (2005).
13) Misu, T. and Kawahara, T.: A bootstrapping approach for developing language model of new spoken dialogue systems by selecting Web texts, *Proc. Interspeech*, pp.9–12 (2006).
14) Steinberger, J. and Ježek, K.: Using Latent Semantic Analysis in text summarization and summary evaluation, *Proc. ISIM*, pp.93–100 (2004).
15) Komatani, K., Fukubayashi, Y., Ogata, T. and Okuno, H.G.: Introducing Utterance Verification in Spoken Dialogue System to Improve Dynamic Help Generation for Novice Users, *Proc. SIGDial*, pp.202–205 (2007).
16) Kawahara, T., Lee, A., Takeda, K., Itou, K. and Shikano, K.: Recent Progress of Open-source LVCSR Engine Julius and Japanese Model Repository, *Proc. ICSLP*, pp.3069–3072 (2004).
17) Chu-Carroll, J. and Carpenter, B.: Dialogue Management in Vector-Based Call Routing, *Proc. COLING-ACL98*, pp.256–262 (1998).
18) Quinlan, J.R.: *C4.5: Programs for Machine Learning.*, Morgan Kaufmann, San Mateo, CA (1993). http://www.rulequest.com/see5-info.html

(平成 20 年 6 月 4 日受付)
(平成 20 年 11 月 5 日採録)

池田　智志（学生会員）

2007 年京都大学工学部情報学科卒業．現在，同大学院情報学研究科知能情報学専攻修士課程在学中．音声対話システムの研究に従事．

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．同年京都大学情報学研究科助手．2007 年より助教，現在に至る．京都大学博士（情報学）．音声対話システムの研究に従事．2008 年から 2009 年まで米国カーネギーメロン大学客員研究員．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等受賞．電子情報通信学会，言語処理学会，人工知能学会，ACL，ISCA 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（現・准教授）．博士（工学）．この間，2005 年より早稲田大学ヒューマノイド研究所客員准教授，2006 年より理化学研究所脳科学総合研究センター客員研究員を兼務．研究分野は人工神経回路モデルおよび人間とロボットのコミュニケーション発達を考えるインタラクション創発システム情報学．2001 年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，SICE，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，科学技術振興事業団，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-2001，2006 Best Paper Nomination Finalist，IROS-2008 Award for Entertainment Robots and Systems Nomination Finalist 2 件，第 2 回船井情報科学振興賞等受賞．人工知能学会，日本ロボット学会，日本ソフトウエア科学会，ACM，IEEE，AAAI，ASA 等各会員．