

マルチドメイン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 ドメイン選択の高精度化

池田 智志[†] 駒谷 和範[†] 尾形 哲也[†] 奥乃 博[†]

[†] 京都大学大学院 情報学研究科 知能情報学専攻

E-mail: †{sikedak,komatani,ogata,okuno}@kuis.kyoto-u.ac.jp

あらまし 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。トピック推定結果は、想定外発話に頑健であるが文脈情報を含まない。一方で、対話履歴は、想定外発話に頑健でないが文脈情報を含むため、これら2つは相補的に働く。話者10名2191発話を用いた評価実験により、従来手法からドメイン選択誤りを14.3%削減した。

キーワード マルチドメイン音声対話システム、ドメイン選択、トピック推定、対話履歴

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

Satoshi IKEDA[†], Kazunori KOMATANI[†], Tetsuya OGATA[†], and Hiroshi G. OKUNO[†]

[†] Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

E-mail: †{sikedak,komatani,ogata,okuno}@kuis.kyoto-u.ac.jp

Abstract 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). 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 2191 utterances showed that our integrated method reduced domain selection errors by 14.3%.

Key words multi-domain spoken dialogue system, domain selection, topic estimation, dialogue history

1. はじめに

事前教示を受けていない一般ユーザが、電話などのインターフェースを通して音声対話システムを使用する状況が増加している。このような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全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る。本論文では、マルチドメインシ

ステム内のサブシステムをドメインと定義する。マルチドメインシステムでは、各ドメインが独立に設計されているため、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。我々は以前に、対話履歴と各ドメインの言語理解結果を利用したドメイン選択手法を開発した[3]。ところが、ユーザの発話が想定外発話であった場合、正しい言語理解結果が得られないため、ユーザの意図した具体的なドメインを推定できない場合があるという問題があった。

本研究では、以下の2つのアプローチにより、この問題に対処する。

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

表 1 トピック推定結果と対話履歴の関係

	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	×	○

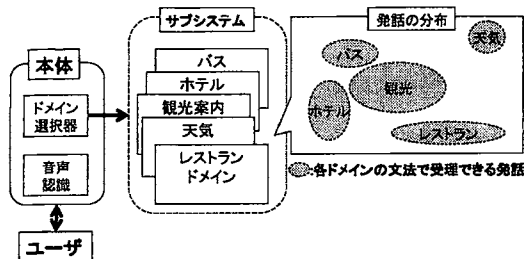


図 1 マルチドメイン音声対話システムのアーキテクチャ

(2) トピック推定と対話履歴との統合 (4. 章に対応)

『ユーザー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。本論文では、トピック推定と対話履歴を統合することで、想定外発話に頑健なドメイン選択器を構築する。トピック推定と対話履歴の関係を表 1 に示す。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、対話履歴は文脈を考慮している。このように、トピック推定と対話履歴は相補的な情報であり、これらを統合することで、より高精度なドメイン選択が可能となる。

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、ユーザーの多様な要求に単一インターフェースで応答できるため、ユーザーにとって利便性が高い。一方で、構築に多大な労力がかかるという問題がある。そのため、ドメインの追加や修正が可能のように (ドメインの拡張性)、個々のドメインを独立に統合してシステムを構築するアーキテクチャが提案されている [2]。システムは複数のドメインとそれらを統合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、ユーザー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。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を図 1 に示す。

2.2 想定外発話への対処のためのトピック

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、正しい言語理解結果が得られず、ドメイン選択誤りを引き起こす。本論文では、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。我々が以前開発

レストランを意図したがシステムは受理できない

例:「個室のある静かな感じの和食のお店」

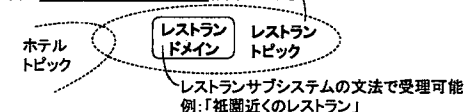


図 2 ドメインとトピックの関係及びその具体例

したドメイン選択手法 [3] も、対話履歴と発話の言語理解結果のみに基づいていたため、想定外発話がドメイン選択誤りの原因となる場合があった。

本研究では、想定外発話への対処として、ユーザーの意図推定を行う。具体的には、システム想定外であっても、あるドメインの内容を意図した発話の集合を“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及びその具体例を図 2 に示す。

トピック推定に関する関連研究として、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 [4]。しかし、発話の話題推定の際にあらかじめ収集されたコーパスを用いた学習をしているため、コーパスのないドメインの学習データの収集が容易ではない。これは、マルチドメイン音声対話システムの構築において重要である、ドメイン拡張性を満たしていない。そこで、本研究ではドメインの拡張性を損なわずにトピック推定を行う。

2.3 システム想定外発話に頑健なドメイン選択

我々は以前、(I) ひとつ前に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) その他のドメインを判別するドメイン選択手法を開発した [3]。ドメイン選択に関する従来研究 [2], [5] では、(I), (II) のみを考慮していた。文献 [3] の手法の問題点として、想定外発話に対して正しい情報が得られず、ユーザーの意図した具体的なドメインを選択できない場合があった。この対話例を図 3 に示す。U2 でユーザー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、一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 (手法 [3]) では、(III) (= (I), (II) のいずれでもない) という正しい判別結果を得ているが、具体的なドメインがわからず、具体的な応答ができない。

そこで本研究では、文献 [3] の手法にトピック推定から得られる情報を加える。本研究のドメイン選択の概略を図 4 に示す。すなわち、(I) ひとつ前の応答を行ったドメイン、(II) 言語理解用音声認識器の認識結果の言語理解に対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、の判別を行う^(注 1)。図 3 の S2 (本手法) のように、トピック推定を用

(注 1): (IV) その他のドメインを定義しているのは、トピック推定を用いても意図に応答すべきドメインを決定できない場合があるからである。例えば、一つ前のユーザー発話においてドメイン選択誤りが生じたため、システムの応答に対して

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 (従来手法): 京都外大前から競馬場までですか? (ドメイン: バス)
 S2 (手法 [3]): 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。例えば、「祇園周辺の寺を検索」などとおっしゃって下さい。 (ドメイン: 観光)

図 3 システム想定外発話を含む対話例

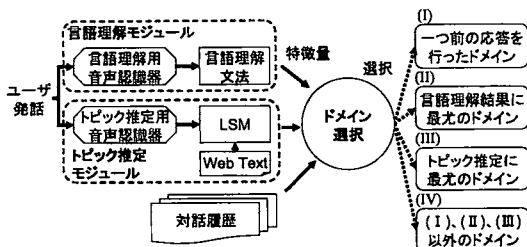


図 4 ドメイン選択の概略

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することができる。想定外発話に対しては言語理解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われる [6]。トピック推定の概略を図 4 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、5 の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「ヘルプ」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 つのアプローチにより行われる。詳細は文献 [6] に示す。

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、ツール [7] を用いて Web から文書を収集した。このツールは、人手で指定した 10 程度のキーワードを用いて収集した Web 文書に対して、

ユーザが「いいえ」と言った場合がそれにあたる (図 5 の U4)。この場合、(I) は誤りであり、(II) と (III) は「いいえ」という発話のみから行われるため、応答すべきドメインを一意に決定することができない。

表 2 ひとつ前の応答を行ったドメインに関する特徴量 [3]

P1:	そのドメインに移移した後のユーザの肯定応答回数
P2:	そのドメインに移移した後のユーザの否定応答回数
P3:	そのドメインに移移する前に、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:	そのドメインに移移する前に、同じドメインであったことがあるか。
P5:	そのドメインに移移してから現在までに変化したスロット数
P6:	そのドメインに移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 (=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

Wikipedia^(注2)から人手で収集した数百文の文書によるフィルタリングを行い、トピックごとに 10 万文を収集する。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。以上の作業で各ドメインごとに収集した文書をランダムに d 個に分割し、学習文書を構成した。

3.2 LSM を用いたトピック推定

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。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、学習文書の k 次元空間でのベクトルを得る。本研究で作成した共起行列は、 $M = 67533$, $N = 120$, $n = 6$, $d = 20$ である。次元縮約に関しては $k = 50$ とした。

トピックと入力発話の近さを、トピックに属する d 個の学習文書の k 次元ベクトルと、入力発話の認識結果の k 次元ベクトルとのコサイン距離の最大値と定義する。ここで、入力発話の音声認識には、Web から収集した文書から構築した統計的言語モデル (トピック推定用言語モデル) を用いる。入力発話に最も近いトピックが、トピック推定結果として出力される。

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、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、(I) ひとつ前の応答を行ったドメイン、(II) 言語理解用音声認識器の認識結果の言語理解に対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、を選択する判別器を、対話データから学習する。以下では、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、文献 [3] で使用した特徴量 (表 2, 3, 4) に加

(注2): <http://ja.wikipedia.org/>

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 [3]

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 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 [3]

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

えて、トピック推定に関する特徴量 (表 5) 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。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。次に、ラベル (I), (II), (III) の関係を表すために、T7~T9 を導入した。例えば、トピック推定で最尤のドメインと一つ前のドメインが一致する場合は、一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報を表す [8]。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

5.1 評価対象の対話データ

評価データとして文献 [3] で収集された対話データを用いる。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、ひとつ前の応答を行ったドメインには、音響尤度に 40

表 5 トピック推定に関する特徴量

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)

表 6 特徴選択の結果得られた特徴量

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	P1, P4, P5, P8, P9, P10, U1, U2, U3, U5, U6, U7, U9, C8, C9, C11

加算して比較した [3]。

言語理解用音声認識には Julian [9] 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。語彙サイズは 7,373 であった。トピック推定用音声認識には、Julius [9] 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。語彙サイズは 56,453 であった。音響モデルは 3000 状態不特定話者 PTM トライフォンモデル [9] を用いた。また、言語理解用音声認識器による単語正解率は 63.3% であり、トピック推定用音声認識器による単語正解率は 69.6% であった。

決定木の構築には、C5.0 [10] 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 2, 3, 4, 5 に示した 43 の特徴量から選択を行った。選択された特徴量を表 6 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ひとつのドメインを選択して正解判定を行った。

発話データは、以下に従って人手でラベル付けされている。

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

ラベル (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

ラベル (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

ラベル (IV) : その他の場合

5.2 ドメイン選択精度の評価

まず、本手法におけるドメイン選択精度を示す。表 7 は、本

表 7 本手法でのドメイン選択の Confusion Matrix

正解 \ 識別結果	ひとつ前 (I)	言語理解最尤 (II)	トピック推定最尤 (III)	その他 (IV)	計 (再現率)
ひとつ前 (I)	1348	34	23	37	1442 (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)	1652 (81.6%)	320 (80.6%)	87 (42.5%)	132 (63.6%)	2191 (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

表 8 ベースラインと本手法の比較 (正解数/発話数)

手法 \ 正解ラベル	一つ前 (I)	言語理解最尤 (II)	その他 (IV)	計
本手法	1348/1442	258/380	140/369	1672/2191
ベースライン	1303/1442	238/380	131/369	1746/2191

手法における正解ラベルと識別結果の Confusion Matrix である。本手法におけるドメイン選択誤り数は 464 となった。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。例えば、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。

次に、以下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: 文献 [3] での提案手法によりドメイン選択を行う。5.1 節の対話データのラベル (III) を (IV) に変更した後、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 2, 3, 4 に示した特徴量から選択した。選択された特徴量を表 6 下段に示す。

ベースライン手法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% (=519/2191) である。ここで、ベースライン手法におけるドメイン選択の正解基準を本手法のドメイン選択に適用した場合、本手法のドメイン選択誤り数は、445 となり、ドメイン選択誤り率は 20.3% (=445/2191) である。ドメイン選択誤り削減率は 14.3% (=74/519) となる。正解ラベルごとのドメイン選択の正解数を表 8 に示す。表 8 によると、全ての正解ラベルにおいて、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。また、表 8 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、ベースライン手法において 1541(=1303 + 238) である。一方で、表 7 によると、本手法では 1643 (=1348 + 258 + 37) であり、ベースライン手法から 102 発話増加している。これは、ベースライン手法より本手法の方が、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。

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

表 9 各特徴量を除いた場合のドメイン選択誤りの増加数

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

特徴量とその増加数を表 9 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。

6. ドメイン選択後の対話戦略

以下では、ドメイン選択後のシステムの動作について述べる。選択されたドメインごとの対話戦略の概要を表 10 に示す。

まず、(I), (II), (III) が選択された場合は、応答すべきドメインが一意に決定されるので、当該ドメインに応じた応答を行う。ユーザ発話が想定内だった場合は、言語理解結果を受理し、タスクを進行する。一方、想定外だった場合は、言語理解結果が信頼できないため、言語理解結果を受理せず、ヘルプを提示する。ヘルプの提示手法には、トピック推定の信頼度に応じたヘルプ提示手法 [11] などが利用できる。また、想定外発話と想定内発話の判定には、音響尤度差に基づく発話検証手法 [8] が利用できる。

次に、(IV) が選択された場合は、一つ前で応答したドメインへの遷移を禁止した上で、一つ前のユーザ発話のドメイン選択をやりなおし、選択されたドメインに応じたヘルプの提示を行う。具体的には、一つ前のユーザ発話を再びドメイン選択器へ入力し、遷移が禁止されたドメイン以外で最もスコアの高いドメインを出力とする。以上の動作を具体的なドメインが決定されるまで繰り返す。これは、(IV) が対話履歴に誤りがある状態を検出していることに着目している。実際、(IV) と正しく選択された 84 発話のうち 83 発話は、一つ前の発話でドメイン選択誤りをしていた。

本論文で述べたドメイン選択手法により、(IV) が選択された場合の対話例を図 5 に示す。この対話例は文献 [3] で収集された対話の一部であり、U1, U2 は、データ収集時のシステムによりドメインが選択されている。まず、U2 において、音声認識誤りのために誤ったドメイン (レストランドメイン) が選択された。ここで、U3 において、従来手法でドメイン選択をし

表 10 ドメイン選択後の動作

ドメイン \ 発話検証	想定内発話	想定外発話
ひとつ前 (I) 言語理解最尤 (II) トピック推定 (III)	言語理解結果を 受理し、タスク進行 ヘルプ生成	言語理解結果を棄却し、 ドメインに応じた ヘルプ生成
その他 (IV)	誤ったドメインへの遷移を禁止したうえで、 一つ前の発話を再びドメイン選択し、 選択結果に応じたヘルプ生成	

U1: 京都駅の宿泊場所を教えてください (ドメイン: ホテル)
S1: 住所が、京都駅前、で宿泊施設を検索します... (○ドメイン: ホテル)
U2: 予算 上限予算 イチマン 円 (ドメイン: ホテル)
(言い淀みのため、「餃子のお店で上限予算一万円」と誤認識)
S2: 上限予算が 10000 円でレストランを検索しますか? (×ドメイン: レストラン)
U3: いいえ
(従来手法: (I) 一つ前のドメインを選択)
(本手法: (IV) その他のドメインを選択)
S3 (従来手法): やり直します。レストランについて、場所、フードタイプなどを指定してください。 (×ドメイン: レストラン)
S3 (本手法): やり直します。ホテルについては、場所、ルームタイプなどを指定してください。 (○ドメイン: ホテル)

図 5 (IV) が選択された場合の対話例

た場合、S3 (従来手法) でシステムは一つ前に応答したドメイン (レストランドメイン) で対話を継続してしまい、誤ったドメインを選択し続けてしまう。一方、本手法では、U3 のドメイン選択結果が (IV) となり、S2 で応答したレストランドメインが誤りであることが検出された。これにより、U2 に対してレストランドメインへの遷移を禁止した上でドメイン選択を行うことで、(I): ホテルドメインが出力される。よって、S3 (本手法) では、正しいドメイン (ホテルドメイン) に対するヘルプの提示を行い、ユーザに再発話を促すことが可能となる。

実際、(IV) が正しく選択された 84 発話のうち、上述の戦略により、58 発話に対して正しいドメインが得られることを確認した。このように、(IV) が検出された場合には、以前のドメイン選択結果に誤りがあるとみなして、ユーザの発話履歴を逆のぼって再度ドメイン選択を行うことで、応答すべきドメインを正しく選択することができる。

7. 結 論

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。対話履歴の利用とトピック推定という相補的な情報を決定木を用いて統合することによって、システムの言語理解可能な範囲を越えた発話に対処することが可能となった。10 名の被験者から収集した対話データ [3] 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%削減された。改善された発話の中には、従来手法では本

質的に扱えなかった発話が含まれていた。また、ドメイン選択の後、特に (IV) が選択された場合の対話戦略について考察した。誤ったドメインへの遷移を禁止した上で、一つ前のドメイン選択をやり直すというアプローチにより、(IV) と正しく推定された 84 発話のうち、58 発話において正しいドメインが選択された。今後は、本論文において開発したドメイン選択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 [7] を用いた。また、評価用対話データは、ホンダ・リサーチ・インスティテュート・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。ここに記して、感謝の意を表す。

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。

文 献

- [1] J. R. Bellegarda: "Latent semantic mapping.", IEEE Signal Processing Mag., 22, 5, pp. 70-80 (2005).
- [2] B. Lin, H. Wang and L. Lee: "A distributed agent architecture for intelligent multi-domain spoken dialogue systems", Proc. ASRU (1999).
- [3] 神田, 駒谷, 中野, 中塚, 辻野, 尾形, 奥乃: "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, 48, 5, pp. 1980-1989 (2007).
- [4] I. R. Lane, T. Kawahara, T. Matsui and S. Nakamura: "Topic classification and verification modeling for out-of-domain utterance detection", Proc. ICSLP, pp. 2197-2200 (2004).
- [5] I. O'Neill, P. Hanna, X. Liu and M. McTear: "Cross domain dialogue modelling: An object-based approach", Proc. ICSLP, Vol. I (2004).
- [6] S. Ikeda, K. Komatani, T. Ogata and H. G. Okuno: "Topic estimation with domain extensibility for guiding user's out-of-grammar utterance in multi-domain spoken dialogue systems", Proc. ICSLP, pp. 2561-2564 (2007).
- [7] T. Misu and T. Kawahara: "A bootstrapping approach for developing language model of new spoken dialogue systems by selecting Web texts", Proc. Interspeech, pp. 9-12 (2006).
- [8] K. Komatani, Y. Fukubayashi, T. Ogata and H. G. Okuno: "Introducing utterance verification in spoken dialogue system to improve dynamic help generation for novice users", Proc. SIGDial, pp. 202-205 (2007).
- [9] T. Kawahara, A. Lee, K. Takeda, K. Itou and K. Shikano: "Recent progress of open-source LVCSR engine Julius and Japanese model repository", Proc. ICSLP, pp. 3069-3072 (2004).
- [10] J. R. Quinlan: "C4.5: Programs for Machine Learning.", Morgan Kaufmann, San Mateo, CA (1993). <http://www.rulequest.com/see5-info.html>.
- [11] 池田, 駒谷, 尾形, 奥乃: "ドメイン拡張性を備えたトピック推定に基づく発話誘導を行うマルチドメイン音声対話システム", SIG-SLUD-A701-10, pp. 83-88 (2007).