Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

ドメイン外発話が扱え拡張性が高い対話ドメイン選択フレームワーク

佐藤 隼[†1] 中野 幹生[†2] 駒谷 和範[†3]
船越 孝太郎[†2] 奥乃 博[†4]

分散型アーキテクチャに基づくマルチドメイン対話システムにおいて，ドメイン選択は非常に重要である．誤ったドメイン選択は対話の進行を著しく阻害するからである．これまでに提案された高精度なドメイン選択法は，特定のタイプのドメインエキスパートでのみ扱える特徴量を使用していた．そこで我々は，2 段階ドメイン選択フレームワークを提案した．この方法では，アクティベートされているエキスパートが，入力発話に対し，継続してその発話を扱うか否かを決定する．このプロセスでは，他のエキスパートがアクティベートされる確率値を考慮する．本稿では，このフレームワークをどのドメインにも属さない発話を扱えるように拡張し，実験結果によりその有効性を示す．

An Extensible Domain Selection Framework That can Deal with Out-of-Domain Utterances

Shun Sato ,[†1] Mikio Nakano ,[†2]
Kazunori Komatani ,[†3] Kotaro Funakoshi [†2]
and Hiroshi G. Okuno [†4]

Domain selection in multi-domain spoken dialogue systems, which employ distributed domain experts, is a crucial problem because an erroneous domain selection disrupts a dialogue. Previously proposed accurate domain selection methods use information available only with specific types of domain experts. We have proposed a two-stage domain selection framework. It decides whether the activated expert should continue to be activated to deal with the input utterance or not. In this process, the probability that another expert is newly activated is used. This paper presents an extension to this framework for dealing with utterances that are not in any domains. It also shows experimental results that show its viability.

1. はじめに

対話インターフェースが広く利用されるようになり，複数のドメインの対話を扱うことができることが期待されている．特に，オフィスロボット[1)] や案内システム[3)] などの音声インタフェースは，単一のドメインしか扱えない特定業務に従事するように自動化されたコールセンターシステムなどとは異なり, 人間の様々な要求に対応できることが期待されている．

そのようなシステムを構築する 1 つの効果的な方法として, 複数の小さなドメインのシステムを統合する**分散型マルチドメイン対話システムアーキテクチャ**がある．このアーキテクチャは，対話状態と音声理解・発話生成のための知識を独立に管理する分散モジュールを用いる（例えば，11)）．分散型アーキテクチャは，ドメインエキスパートを他のドメインエキスパートと独立に設計でき，新しいドメインエキスパートを加えるのが容易であるという工学的利点がある．各ドメインの対話戦略は他のドメインの対話戦略と全く異なっていても良い．例えば，フレームベースの混合主導の対話管理，有限状態ベースのシステム主導の対話管理，プランベースの対話管理などを用いることができる[12)]．

分散型アーキテクチャの重要な課題の一つは，各ユーザ発話に対して，適切なドメインを選ぶことである．今までに，様々な方法が提案されたが，それらは，次の二つの条件を同時には満たさない．一つは，様々なタイプのドメインエキスパートが使用できること (**拡張性**) であり，もうひとつは，音声認識誤りに対して頑健であること (**頑健性**) である．これが，マルチドメイン対話システムが有用だと認識されているのにも関わらず，あまり開発されていない原因の一つであると考える．

我々は，上記の二つの条件を満たす，2 段階ドメイン選択フレームワークを提案した[19)]．このフレームワークは，各エキスパートが二つのサブモジュールを持つと仮定する．現在アクティベートされていない場合に入力発話によりアクティベートされる確率を推定するモ

†1 東京電機大学 大学院理工学研究科
Graduate School of Science and Engineering, Tokyo Denki University
†2 (株) 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.
†3 名古屋大学 大学院工学研究科
Graduate School of Engineering, Nagoya University
†4 京都大学 大学院情報学研究科
Graduate School of Informatics, Kyoto University

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

ジュールと, すでにアクティベートされているとき，継続されるかどうか判定するモジュールである．ドメイン選択のために各エキスパートで行うプロセスを他のエキスパートとは独立に設計できるため，拡張性を満たす．また，これらのサブモジュールは, 対話履歴と音声理解に関するドメイン依存の情報を用いることができるため，頑健性を満たすことができる．特に，対話履歴を用いることで，誤ったドメイン遷移を避けることができる．

しかしながら，この方法では，どのドメインにも属さない発話をどう扱うかが不明であった．そこで我々は，2 段階ドメイン選択法を拡張し，どのドメインにも属さない発話を扱えるようにした．

本稿では，対話履歴を使わないで, 各発話を分類する問題 (例えば, 2)) は扱わない．対話履歴を考慮して，ドメインが継続するか行するかのユーザーの意図を推定することに主眼をおいている．

2. マルチドメイン対話システムのドメイン選択

2.1 分散型アーキテクチャ

分散型マルチドメイアーキテクチャでは，各モジュールが独自に発話理解と，発話理解の知識を保持している[11),13),17),18)]. 我々は，これらのモジュールのことを**ドメインエキスパート**と呼ぶ．分散型マルチドメイン対話アーキテクチャ(図 1) では入力発話の音声認識結果がそれぞれのエキスパートにとして送られる．エキスパートは，自分の持っている発話理解用知識を用いて，その音声認識結果の理解を試みる．ドメイン選択部はそれぞれのエキスパートから情報を得て，どのエキスパートがその発話を理解してシステム応答を行うべきかを決定する．本稿では，各時点で選択されて発話理解・生成を行っているエキスパートを**アクティベート**されていると呼ぶ.

2.2 システムの例

今までに，分散型アーキテクチャに基づいて，多くのマルチドメイン対話システムが開発され，様々な状況で対話出来ることを示している．例えば，複数のドメインでの情報提供やデータベース検索を行うシステム[3),8),11),16)] や， 異なった対話戦略を用いる複数のドメインを統合したものなどがある．例えば，タスク指向と非タスク指向の対話を統合したシステム[10),14)] や，対話だけでなく物理動作を用いるドメインエキスパートも用いるシステムがある[13)].

以下に，5 節の評価実験で用いた対話データを集めるのに用いた対話システムを説明する．

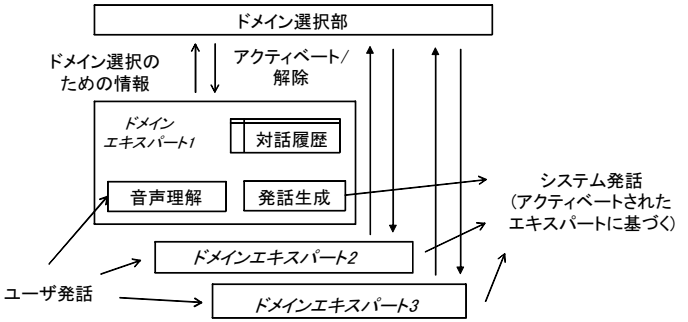

図 1 分散型マルチドメインアーキテクチャ

スライドと Microsoft Agent[⋆1] を用い，ユネスコ世界遺産を紹介するシステムで，以下の 10 個のドメインエキスパートを持つ.

質問応答エキスパート (QA): このエキスパートは世界遺産に関する質問応答ペアを持つデータベースを用いる[15)]. 各質問応答ペアは，いくつかの質問例と，1 ないし数文およびスライドからなる応答からなる．それぞれの質問例は，キーフレーズがマークされており，入力発話と質問例のマッチングに使われる.

インタラクティブプレゼンテーションエキスパート (IP): 有名な世界遺産の 1 つを詳しく説明するエキスパートが 8 つあり， ユーザが説明を中断した時，システムは割り込み発話に応じた内容の説明をする[15)]．例えば，説明を付け加えたり説明を止めたりすることができる．割り込み発話のパターンは IP エキスパートの知識ベースに定義されている.

要求理解エキスパート (RU)：このエキスパートは上で説明したインタラクティブプレゼンテーションエキスパートの 1 つを開始する要求を理解し，その要求を確認するための対話を行う．理解が終わると，理解した要求をタスク計画部と呼ばれるモジュールに送る．そして，タスク計画部は要求されたプレゼンテーションを行う IP エキスパートをアクティベートする.

このシステムのドメイン選択は，手動で構築した規則を用いた．図 2 に人間とこのシステムとの対話例を示す．ユーザー発話は比較的短く，特定の世界遺産か地域の単語を含んでいる．それらの単語の認識に失敗すると，対話履歴が用いられていない場合に正しくドメイン選択を行うことが難しい.

⋆1 http://www.microsoft.com/MSAgent/

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

	発話	エキスパート
S1:	こんにちは，イタリア，ドイツ，ギリシャと他の地域の有名な世界遺産についての質問にお答えします．また，いくつかの世界遺産については，詳しく説明できます．何でも聞いてみてください．	QA
U1:	イタリアの世界遺産のリストを見せてください．	QA
S2:	(リストを表示) イタリアの世界遺産を地域別に紹介します．どの地域のリストが見たいかおっしゃって下さい．	QA
U2:	他に何が聞けますか？	QA
S3:	(リストを表示) ドイツ、イタリア、ギリシャの世界遺産と，他の著名な世界遺産に関して質問にお答えします．	QA
U3:	グリニッジについて教えて下さい	RU
S4:	海事都市グリニッジの説明ですね？	RU
U4:	はい．	RU
S5:	(グリニッジのスライドを表示) はい． 海事都市グリニッジの案内をします．グリニッジ標準時の基準となる都市として由緒ある港町です．	IP3
U5：	グリニッジ標準時ってなんですか？	IP3
S6:	(グリニッジ標準時に関するスライドを表示) グリニッジ天文台での平均太陽時のことです．	IP3

図 2 データ収集用システムの対話例．IP3 はグリニッジを説明するインタラクティブプレゼンテーションエキスパート．S はシステム発話，U はユーザ発話を示す．

2.3 ドメイン選択法が満たすべき条件

分散型アーキテクチャのドメイン選択は，以下の 2 つの条件を満たす必要がある．

拡張性　分散型アーキテクチャの拡張性を損なわない．すなわち，どのようなエキスパートでも用いることができなければならない．そして，各エキスパートは独立に設計出来なければならない．したがって，ドメイン選択部がエキスパートから得る情報を，どのようなエキスパートでも出力できる単純なものにする必要がある．

頑健性　音声認識誤りに頑健である必要がある．すなわち，誤認識に起因する，誤ったドメイン遷移を防ぐ必要がある．

3. 従来のドメイン選択手法

今までに様々なドメイン選択手法が提案されているが，我々が知る限り，拡張性と頑健性の両方を満たしている方法はない．Isobe らの方法 6) は，音声認識の結果から各ドメインのスコアを見積もり，最も高いスコア (以後 **RecScore** と呼ぶ) を持ったドメインを選択する．これは，それぞれのドメインエキスパートが数値スコアを出力するだけで良いので，拡張性を満たす．しかしながら，この方法は対話文脈を考慮していないので，あるエキスパートのスコアが偶然高くなってしまったとき，ドメインを誤って移行させる傾向がある．例えば，図 2 の U4 で，たまたま **QA エキスパート**のスコアが高くなってしまった時に，ドメイン選択を失敗してしまい，頑健性を満たさない．

誤ったドメイン遷移を防ぐために，直前のドメインと同じドメインのスコアに一定値を加える方法[11] がある (**RecScore+Bias** と呼ぶ)． しかし，ドメインが続くかどうかは対話文脈に依存する．例えば，あるドメインの対話タスクが終わった時には，違うドメインに移行する可能性が高いと考えられる．よって，固定スコアを加えれば，いつも上手くいくとは限らない．O'Neill らの方法[16] は，そのドメインのタスクが終わるまで，システムは対話ドメインを変えられない．この方法は誤ったドメイン遷移から抜けることができないという問題がある．.

音声認識誤りに対する頑健性を満たすため，音声認識結果と対話履歴に関する特徴を使用した分類器を用いてドメインを選択する方法がある[5),8),10)]．しかしこれらの方法は，特定のドメインエキスパートでしか得られない特徴量を用いているため，拡張性を満たさない．

単語 (および n-gram) の出現頻度に基づく分類器を用いる方法が，発話分類[2)]，音声コーパスのトピック推定[4)]，人間同士の対話のデータのトピック推定[9)] に用いられている．これらの方法は，マルチドメイン対話システムにおけるドメイン選択に適用できる．しかしながら，これらの方法は，ターゲットシステムとと同じドメインの学習データを利用しなければいけないので，拡張性は失われる．さらに，対話や音声理解に関係した様々な特徴量を用いることができないので頑健でない．2.2 節で説明したシステムのように，複数のドメインで多くの単語が共有されている場合には，単語出現頻度は必ずしも有効であるとは限らない．

4. 2 段ドメイン選択フレームワーク

4.1 提案フレームワーク

拡張性を満たすため，ドメイン選択部がエキスパートから得る情報を，どのようなエキスパートでも出力できる単純な数値スコアに制限する必要がある．**RecScore** と **RecScore+Bias** はそれを満たす．しかし，上で述べたように，音響スコアに直前のエキスパートの閾値を加える方法は精度が悪い．

精度をあげるための 1 つの拡張法として，音声認識結果のだけでなく，各エキスパートに依存した，対話履歴や音声認識などに関連する特徴を使用することが考えられる．各エキスパートは，入力発話が，どれくらいそのエキスパートに適合しているかの確率値を，特

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

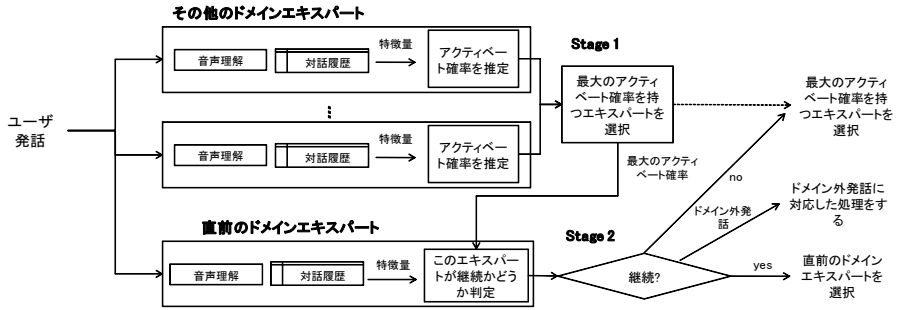

図 3 2 段ドメイン選択フレームワーク

徴量から見積もり，最も高い確率値を出したエキスパートを選択する (**Max-Prob**)．ドメイン選択部が，エキスパートに依存した特徴量を直接使用しないので，この方法は拡張性を満たす．しかしながら，直前のドメインのエキスパート以外のエキスパートが，間違って高い確率値を出力すると, 前のドメインエキスパートの対話状態に関係なく遷移するので, **RecScore**，**RecScore+Bias** と同じ問題がある．

我々は，ドメインが頻繁に遷移しない点に注目した．我々の方法は，ドメインが継続するかどうかをまず判定し，継続しないと判定された場合のみ，どのドメインに遷移するかを判断する．そして，ユーザ発話が，直前のドメインから遷移しないと文脈から判断できる時，誤ったドメイン遷移を防ぐことが出来る．ドメインが非継続と判定されたら，各エキスパートで次にアクティベートされる確率値を計算し，確率値が最も高くなるエキスパートを選択する．

この方法は，入力発話がどれくらいドメインを遷移させるか，を考慮に入れていないという問題がある．そのため，他のエキスパートが，どのくらいアクティベートされるかの確率値を推定し，その最大値を用いる．これは，各エキスパートが出力する唯一の情報である．したがって，拡張性は損なわれない．また，**RecScore**，**RecScore+Bias** と異なり，最大のアクティベート確率が高かったとしても，直前のドメインエキスパートが継続するかどうかは直前のエキスパートの内部状態に基づいて判定することが出来る．従って頑健性を満たす．

このアイディアに基づき，我々は 2 段階ドメイン選択フレームワークを提案した[19]．しかしながら，これは，どんなドメインにも属さない発話 (ドメイン外発話 (out-of-domain utterance) と呼ぶ) を考慮していなかった．

そこで我々は 2 段階ドメイン選択フレームワークを拡張し，ドメイン外発話を扱うようにする．これは，直前のドメインが継続するかどうかの判定の際に，ドメイン外発話かどうかも判定するようにした．すなわち，継続か，非継続か，ドメイン外発話かの 3 カテゴリへの分類を行う．

拡張したフレームワークをまとめると以下のようになる (図 3)．

ステージ 1 では，アクティベートされていないエキスパートが，入力発話によりアクティベートされる確率を推定し，ドメイン選択部に送る．ドメイン選択部はその最大値を求める．ステージ 2 では，直前のドメインのエキスパートが，入力発話を引き続き処理すべきかどうか，またはドメイン外発話かどうかを判定する．この判定では，ステージ 1 で得られたアクティベート確率の最大値を使用する．継続しないと判断した場合，ステージ 1 で最も高い確率値を出したエキスパートのドメインを選択する．

フレームワークという言葉を使う理由は, それぞれのドメインエキスパートのドメイン選択用サブモジュールで使われるアルゴリズムや特徴量を特定せずに, インターフェースのみを指定するためである．**RecScore**, **RecScore+Bias**, **Max-Prob** はこのフレームワークの 1 つの実装とみなすことができるが，このフレームワークは, 様々な特徴量を使用することを可能にし，かつ，ドメイン選択用サブモジュールの柔軟な設計を可能にする．

5. 評 価 実 験

各エキスパートのためのアクティベート確率を計算するサブモジュールと，ドメインが継続するかどうかを判定するサブモジュールが，十分なデータで訓練されているなら，提案フレームワークは，従来法の拡張であるため，高い精度を達成すると推測される．このことと，人間とシステムの対話データを用いた実験が, 提案フレームワークの有効性を示すのに十分だと考える．以下で，その実験を説明する．

5.1 デ ー タ

実装と評価実験のために, 2.2 節で述べた世界遺産説明システムと人間との対話のコーパスを使用した．ドメイン選択は人手で構築した規則を使用した．

35 人 (男 17 人，女 18 人，年齢:19-57) の被験者が，4 回システムと対話をした．対話の仕方に関して特に教示はせず，自由に対話してもらった．1 回の対話は 8 分である．23 人分のデータ (3,530 発話) を学習データとし, さらにデータセット A(1,672 発話), データセット B(1,858 発話) に分けて，それぞれを 2 つのステージの訓練データとした． 残りの 12 人分のデータ (1,865 発話) をテストデータとし，各発話に対して，正解ドメインを与えた．表

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

ドメイン	直前のドメイン	トレーニングデータ A	トレーニングデータ B	テストデータ
RU	RU	134	169	145
	QA	51	102	59
	IP	21	16	23
	小計	206	287	227
QA	RU	46	55	51
	QA	783	870	888
	IP	59	87	66
	小計	888	1,012	1,005
IP	RU	2	1	3
	QA	7	11	18
	IP	311	305	277
	小計	320	317	298
OOD	RU	24	19	39
	QA	168	155	183
	IP	66	68	113
	小計	258	242	335
合計		1,672	1,858	1,865

表 1 学習データとテストデータの各ドメインの発話数

1 に各セットの詳細な数字を載せる．

5.2 実　　装

5.2.1 エキスパートクラス

8 つある IP エキスパートは対話戦略と，予測発話パタンの大部分を共有している．各 IP エキスパートの学習データが少ないため, すべての学習データを合わせて，共通の言語モデル，共通のアクティベーション確率推定器，共通のドメイン継続決定器を用いた．そこで，今後 IP エキスパートをまとめて，IP エキスパートクラスと呼ぶ．RU エキスパート, 質問に答える QA エキスパートもそれぞれエキスパートクラスとする．

5.2.2 発 話 理 解

本システムの音声認識は Julius[7] を用いた．音響モデルは日本語ディクテーションキット[7] に含まれる音響モデル (3000-state Phonetic Tied-Mixture model) を用いた．それぞれのエキスパートの音声理解の特徴を表 2 に示す．言語モデルはデータ収集に使用したシステムのものよりも，訓練データを用いて強化されている．各ドメインのテストデータにおける音声認識の性能を示すため，音素誤り率を求めた[⋆1]．

expert class	QA	IP	RU
LM for ASR	trigram	trigram	有限状態
言語理解	キーフレーズマッチング	キーフレーズマッチング	FST ベース
語彙数	1,140	407	79
音素誤り率 (%)	10.95	19.47	23.60

表 2 各エキスパートの音声理解

5.2.3 ステージ 1

ステージ 1 では，アクティベートされていないエキスパートのアクティベート確率をロジスティック回帰で求めた．音声認識に関するものと対話履歴を特徴量として使用した．特徴量を表 3 に示す．これらの特徴量はエキスパートクラスに依存している．これにより，入力発話がどのくらい文脈に適合しているかを，どんな種類のエキスパートでも得られる特徴量しか用いないのに比べ，より正確に推定することができる．

アクティベート確率推定器を訓練するために，データマイニングツールキット Weka[20] のロジスティック回帰[⋆2]を使用した．訓練データ A で訓練してロジスティック回帰係数を得た．各エキスパートクラスのための学習データは，直前のドメインが自分のクラスでなかった発話の時の発話とした．それは，アクティベート確率を見積もるのは，直前のドメインが自分のクラスでなかった発話のみに対して行われるからである．

各々の発話が，エキスパートクラスのドメインの発話なら，'acitavate' ラベルを割り当て，そうでなければ，'non-activate' ラベルを割り当てる．次に過学習を防ぐため，backward-stepwise 法で，訓練データ B 重み付き F 値 ('activate'，'non-acitvate' ラベルの数で重み付けられたもの) がで最大になるように特徴選択を行った．表 4 に残った特徴とその重要度を載せる．

学習データ A で'activate' の発話と'non-activate' の発話の比を，'activate' 発話の量を増やして 1:3 の割合にした．これは, 学習データが'non-activate' の発話を多く含んでおり, 結果に偏りが出てしまうためである．この比率は, 学習データ B の重み付き F 値が最も高くなるように実験的に求めた．

⋆1 単語誤り率を求めなかったのは，長い世界遺産名などがあり，単語の長さのばらつきが大きいからである．
⋆2 Weka のデフォルト値を用いた．

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

expert class	特徴量	
all classes	$F_{i,r1}$	$\mathrm{SRR}_{i,1}$ が得られたかどうか
$i = ru, ip, qa$	$F_{i,r2}$	$\mathrm{SRR}_{i,1}$ にフィラーが含まれるかどうか
	$F_{i,r3}$	min ($\mathrm{SRR}_{i,1}$ の単語信頼度)
	$F_{i,r4}$	avg ($\mathrm{SRR}_{i,1}$ の単語信頼度)
	$F_{i,r5}$	($\mathrm{SRR}_{i,1}$ の音響スコア) / 発話の長さ (sec)
	$F_{i,r6}$	$\mathrm{SRR}_{i,1}$ の言語スコア
	$F_{i,r7}$	$\mathrm{SRR}_{i,1}$ の単語数
	$F_{i,r8}$	$\mathrm{SRR}_{i,all}$ の単語数
	$F_{i,r9}$	($F_{i,r5}$ - ($\mathrm{SRR}_{lv,1}$) の音響スコア) / 発話の長さ
RU	$F_{ru,r10}$	$\mathrm{SRR}_{ru,1}$ が肯定発話か
	$F_{ru,r11}$	$\mathrm{SRR}_{ru,1}$ が否定発話か
	$F_{ru,r12}$	LM_{ru} の認識結果の候補数
	$F_{ru,r13}$	$\mathrm{SRR}_{ru,1}$ に世界遺産の単語が含まれているか
	$F_{ru,h1}$	$\mathrm{SRR}_{ru,1}$ が肯定発話かどうか (ステージ 2 のみ使用)
	$F_{ru,h2}$	RU エキスパートへの遷移後のターン数
	$F_{ru,h3}$	RU エキスパートへの遷移後の否定発話の回数
	$F_{ru,h4}$	$F_{ru,h4}/F_{ru,h3}$
IP	$F_{ip,r10}$	ユーザーがプレゼンテーションを止めようとしてるかどうか
	$F_{ip,r11}$	$\mathrm{SRR}_{ip,1}$ とマッチする質問例がデータベースにあるか
	$F_{ip,r12}$	$\sum_j$((SRR$_{ip,j}$ のキーフレーズが出現する回数) / ($\mathrm{SRR}_{ip,j}$ の単語数)) / (認識結果の個数)
	$F_{ip,r13}$	$\min_i$($\mathrm{SRR}_{ip,all}$ のキーフレーズ i の全認識候補で出現回数 / (音声認識結果の個数))
	$F_{ip,r14}$	$\max_i$($\mathrm{SRR}_{ip,all}$ のキーフレーズ i の全認識候補で出現回数 / (音声認識結果の個数))
	$F_{ip,r15}$	avg($\mathrm{SRR}_{ip,1}$ のキーフレーズ i の第一候補における単語信頼度)
	$F_{ip,r16}$	$\min_i$ ($\mathrm{SRR}_{ip,1}$ のキーフレーズ i の第一候補における単語信頼度)
	$F_{ip,r17}$	$\max_i$ ($\mathrm{SRR}_{ip,1}$ のキーフレーズ i の第一候補における単語信頼度)
	$F_{ip,h1}$	このエキスパートが既出か否か
QA	$F_{qa,r10}$	$F_{ip,r12}$ と同様
	$F_{qa,r11}$	$F_{ip,r13}$ と同様
	$F_{qa,r12}$	$F_{ip,r14}$ と同様
	$F_{qa,r13}$	$F_{ip,r15}$ と同様
	$F_{qa,r14}$	$F_{ip,r16}$ と同様
	$F_{qa,r15}$	$F_{ip,r17}$ と同様
	$F_{qa,r16}$	$\mathrm{SRR}_{qa,1}$ が相槌リストに含まれるかどうか
	$F_{qa,h1}$	$\mathrm{SRR}_{qa,1}$ が肯定発話かどうか (ステージ 2 のみ使用)
	$F_{qa,h2}$	QA エキスパートに遷移後のターン数
	$F_{qa,h3}$	QA エキスパートに遷移後の否定発話の回数
	$F_{qa,h4}$	$F_{qa,h4}/F_{qa,h3}$

$\mathrm{SRR}_{i,j}$ はエキスパートクラス i の言語モデルを用いた音声認識の j 番目の結果. $\mathrm{SRR}_{i,all}$ は n-best の全ての認識結果. $\mathrm{SRR}_{lv,j}$ は発話検証のために用いた大語彙統計モデル[7] での音声認識結果 (60,250 語). F_{i,r_x} は音声理解に関する特徴. F_{i,h_x} は対話履歴に関する特徴. CM は単語信頼度.

表 3 特徴量

エキスパートクラス (特徴選択を行った後の F_1 値)	残った特徴量 (特徴選択後に特徴量を除いた時の F_1 値)
RU (0.939)	$F_{ru,r9}$ (0.926), $F_{ru,r13}$ (0.931), $F_{ru,r5}$ (0.931), $F_{ru,r14}$ (0.935), $F_{ru,r3}$ (0.936), $F_{ru,r2}$ (0.936), $F_{ru,r10}$ (0.937),$F_{ru,r8}$ (0.937), $F_{ru,r12}$ (0.938),$F_{ru,r11}$ (0.938)
IP (0.735)	$F_{ip,r15}$ (0.660), $F_{ip,r14}$ (0.676), $F_{ip,r8}$ (0.701), $F_{ip,r7}$(0.704), $F_{ip,r12}$ (0.708), $F_{ip,r9}$ (0.708)
QA (0.831)	$F_{qa,r6}$ (0.804), $F_{qa,r13}$ (0.810), $F_{qa,r16}$ (0.817), $F_{qa,r10}$ (0.821), $F_{qa,r7}$ (0.821), $F_{qa,r5}$ (0.827), $F_{qa,r4}$ (0.828), $F_{qa,r15}$ (0.829)

表 4 ステージ 1 で特徴選択後に残った特徴量を, 取り除いた時の F 値.

エキスパートクラス (特徴選択を行った後の F_1 値)	残った特徴量 (特徴選択後に特徴量を除いた時の F_1 値)
RU (0.748)	$F_{ru,a}$ (0.693), $F_{ru,r2}$ (0.714), $F_{ru,r9}$ (0.719), $F_{ru,r3}$ (0.719), $F_{ru,h4}$ (0.728), $F_{ru,r5}$ (0.730), $F_{ru,r12}$ (0.736),$F_{ru,r4}$ (0.737), $F_{ru,r15}$ (0.739), $F_{ru,r7}$ (0.739), $F_{ru,r14}$ (0.740), $F_{ru,h3}$ (0.740), $F_{ru,h1}$ (0.742)
IP (0.823)	$F_{ip,a}$ (0.779), $F_{ip,r6}$ (0.801), $F_{ip,r4}$ (0.802), $F_{ip,r8}$ (0.803), $F_{ip,r10}$ (0.804), $F_{ip,r5}$ (0.804), $F_{ip,r3}$ (0.809), $F_{ip,r15}$ (0.811), $F_{ip,r12}$ (0.814), $F_{ip,r13}$ (0.815), $F_{ip,r2}$ (0.815)
QA (0.876)	$F_{qa,a}$ (0.836), $F_{qa,r5}$ (0.861), $F_{qa,r7}$ (0.867), $F_{qa,r6}$ (0.869), $F_{qa,r9}$ (0.870), $F_{qa,r8}$ (0.871), $F_{qa,r3}$ (0.871), $F_{qa,h1}$ (0.872), $F_{qa,h2}$ (0.874), $F_{qa,r13}$ (0.874), $F_{qa,r2}$ (0.875), $F_{qa,r16}$ (0.875), $F_{qa,h3}$ (0.875), $F_{qa,h4}$ (0.875)

表 5 ステージ 2 で特徴選択後に残った特徴量を, 取り除いた時の F 値

5.2.4 ステージ 2

ステージ 2 では, 継続か非継続かシステム想定外発話かを決定するために SVM(サポートベクトルマシン) を使用した. ⋆1 特徴量は, ステージ 1 で得られた最大アクティベート確率値とステージ 1 で使用した特徴量セットを使用した. 各エキスパートクラスの SVM の学習データは, 訓練データ B の発話のうち, 直前のドメインが, そのクラスのエキスパートのドメインであるような発話である. これらの発話は, 継続, 非継続, ドメイン外のいずれかのラベルがついている. 次に, backward-stepwise 法で訓練データ A の F 値が最大になるように特徴選択を行った. 残った特徴を表 5 に載せる. アクティベート確率の最大値はすべてのエキスパートクラスで最も重要な特徴であることが分かる. この結果は, アクティベート確率の最大値を使用する 2 段フレームワークが有効であることを示している.

⋆1 Weka の SMO で線形カーネルを用いた. パラメータはデフォルト値を用いた.

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

非継続の発話とドメイン外発話が, 継続の発話に比べ, 非常に少ないので, 継続，非継続，ドメイン外発話の数の比が 3:1:1 になるように, 非継続の発話, ドメイン外発話のデータを複製した．この比率は, 学習データ A の重み付き F 値が最も高くなるように実験的に求めた．

5.3 評　　価

5.3.1 各方法の比較

5.2 節で説明した手法 (**FullImpl**) と, 以下の 4 つの方法を比較した．これらの方法はすべて拡張性を満たす．最初の 3 つは 4 節で説明した方法である．

5.3.1.1 RecScore

発話時間によって正規化した音声認識結果の音響スコアが最も高いエキスパートクラスを選択する[6]. IP エキスパートは, システムが IP エキスパートから誤って抜けた時に限り, IP ドメイン以外のエキスパートから IP エキスパートに戻ってくるので，最も最近選ばれた IP エキスパートに遷移するとした．全てのエキスパートの音声認識結果の音響スコアが閾値より低かった場合にドメイン外発話とした．閾値は学習データを使用し, 5.3.2 節と同様に, ドメイン内発話とドメイン外発話の分類の重み付き F 値が最大になるよう実験的に求めた．

5.3.1.2 RecScore+Bias

RecScore とほぼ同じだが，直前のエキスパートの場合，スコアにバイアスを加えるよう拡張した．それぞれのエキスパートのバイアスは，訓練データの重み付き F 値が最大にするように決めた．ドメイン外発話の判定は **RecScore** と同様に行った．

5.3.1.3 Max-Prob

ロジスティック回帰を使用し，全てのエキスパートのアクティベート確率を求め，最大の確率値を持つエキスパートを選択した．システムの制限を考慮し，対話中に一度も出てきていない IP エキスパートは除外した．ドメインの継続性を考慮するため，(**FullImpl** で使っている特徴量に加え，直前のドメインを特徴量として用いた．また，各エキスパートのアクティベート確率を求めるのと同様に，ドメイン外発話である確率の推定も行い，この確率が最大値であれば，ドメイン外発話であるとした．特徴選択も行っている．

5.3.1.4 NoActivProb

FullImpl のステージ 2 の特徴量のうち，ステージ 1 で求めたアクティベート確率値の最大値を用いない

5.3.2 結　　果

ドメインの遷移に注目し，ドメイン選択結果を継続, ドメイン遷移，ドメイン外発話検出の 3 つに分類して評価した．評価基準として，重み付き F 値を使用した．重みは，これらの分類の正解ラベルの比率である．ドメイン遷移の精度を求めるとき, 不正解のドメインに遷移する場合は不正解とする．

方法	分類	recall	precision	F 値	重み付き F 値
RecScore	継続	0.763	0.867	0.812	
	遷移	0.559	0.239	0.335	
	ドメイン外	0.501	0.848	0.630	0.789
RecScore +Bias	継続	0.917	0.824	0.868	
	遷移	0.400	0.421	0.410	
	ドメイン外	0.501	0.848	0.630	0.838
MaxProb	継続	0.856	0.877	0.866	
	遷移	0.382	0.238	0.293	
	ドメイン外	0.316	0.453	0.373	0.823
NoActiv Prob	継続	0.883	0.875	0.879	
	遷移	0.377	0.411	0.393	
	ドメイン外	0.830	0.818	0.824	0.844
FullImpl	継続	0.913	0.905	0.909	
	遷移	0.541	0.589	0.564	
	ドメイン外	0.854	0.839	0.846	**0.885**

表 6 評価結果

表 6 に結果を示す．2 項検定によりすべての手法の差が統計的に有意であることが分かった ($p < .01$)

5.3.3 考　　察

FullImpl が, 他の方法より優れている理由の 1 つにドメイン遷移の精度が優れていることがあげられる．それは誤ったドメイン遷移を防ぐことを意味しているので，このフレームワークは頑健であるといえる．**RecScore+Bias** は限られた特徴だけを使用しているが, 比較的精度が良い．用いたデータではドメイン遷移が少ないため，バイアスが有効であったと考えられる．しかしながら，ドメイン外発話に対する F 値が低いことから，認識スコアだけでは不十分であることを示している．よって, ドメイン遷移には, 前のドメインの特徴を使うことが有効であると推測する．**FullImpl** と **NoActivProb** の比較から, ステージ 2 で最大アクティベート確率を用いるのは有効であるといえる．

FullImpl でも, 他の方法よりは高いものの, ドメイン遷移の F 値は低い．典型的な理由の一つは，ある発話の音声認識結果の中のキーワードの一つが直前のエキスパートのドメインの語彙にもあれば，間違って直前のドメインが継続すると判定されやすくなることである．例えば，「他の世界遺産について教えてください」は，正解は QA ドメインだが, IP ドメインの語彙に「世界遺産」が含まれているので, IP ドメインから QA ドメインへの遷移

Vol.2011-NL-201 No.12
Vol.2011-SLP-86 No.12
2011/5/17

が上手く行われない場合がある．これは，ドメイン遷移する場合の学習データが十分でないためだと考えられるため，より多くの訓練データを用いることで解決できると推測できる．

6. 終わりに

本稿では，以前提案した拡張性の高いマルチドメイン対話システムのためのドメイン選択のフレームワークを，ドメイン外発話が扱えるように拡張したものを提案した．このフレームワークは，柔軟な特徴量の利用やドメインの継続性の考慮を可能にすることで，頑健なドメイン選択器を構築することが可能である．このフレームワークにより，マルチドメインの対話システムや会話ロボットなどの構築が盛んになると期待している．

今後の課題として，高いドメイン精度を達成するには，それぞれのドメインエキスパートのアクティベート確率推定やドメイン継続判定が，どのくらい正確であればいいのかを調べることがあげられる．また，このフレームワークのスケーラビリティについて確かめるため，より多くのエキスパートを持つシステムで実験を行う．また，ドメイン選択の確信度を推定する方法も必要であると考えている．

参考文献

1) H.Asoh, T.Matsui, J.Fry, F.Asano, and S.Hayamizu. A spoken dialog system for a mobile office robot. In *Proc. 6th Eurospeech*, pp. 1139–1142, 1999.
2) J. Chu-Carroll and B. Carpenter. Vector-based natural language call routing. *Comp. Ling.*, 25(3):361–388, 1999.
3) J.Gustafson and L.Bell. Speech technology on trial: Experiences from the August system. *Natural Language Engineering*, 6(3&4):273–286, 2000.
4) A.Heidel and L.Lee. Robust toipc inference for latent semantic language model adaptation. In *Proc. ASRU-07*, pp. 177–182, 2007.
5) S.Ikeda, K.Komatani, T.Ogata, and H.G. Okuno. Extensibility verification of robust domain selection against out-of-grammar utterances in multi-domain spoken dialogue system. In *Proc. Interspeech-2008 (ICSLP)*, pp. 487–490, 2008.
6) T.Isobe, S.Hayakawa, H.Murao, T.Mizutani, K.Takeda, and F.Itakura. A study on domain recognition of spoken dialogue systems. In *Proc. Eurospeech-2003*, pp. 1889–1892, 2003.
7) T.Kawahara, A.Lee, K.Takeda, K.Itou, and K.Shikano. Recent progress of open-source LVCSR engine Julius and Japanese model repository. In *Proc. Interspeech-2004 (ICSLP)*, pp. 3069–3072, 2004.
8) K.Komatani, N.Kanda, M.Nakano, K.Nakadai, H.Tsujino, T.Ogata, and H.G. Okuno. Multi-domain spoken dialogue system with extensibility and robustness against speech recognition errors. In *Proc. 7th SIGdial Workshop*, pp. 9–17, 2006.
9) I.R. Lane and T.Kawahara. Incorporating dialogue context and topic clustering in out-of-domain detection. In *Proc. ICASSP-2005*, pp. 1045–1048, 2005.
10) C.Lee, S.Jung, S.Kim, and G.G. Lee. Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, 51(5):466–484, 2009.
11) B.Lin, H.Wang, and L.Lee. A distributed architecture for cooperative spoken dialogue agents with coherent dialogue state and history. In *Proc. ASRU-99*, 1999.
12) M.F. McTear. *Spoken Dialogue Technology*. Springer, 2004.
13) M.Nakano, K.Funakoshi, Y.Hasegawa, and H.Tsujino. A framework for building conversational agents based on a multi-expert model. In *Proc. 9th SIGdial Workshop*, pp. 88–91, 2008.
14) M.Nakano, A.Hoshino, J.Takeuchi, Y.Hasegawa, T.Torii, K.Nakadai, K.Kato, and H.Tsujino. A robot that can engage in both task-oriented and non-task-oriented dialogues. In *Proc. Humanoids-2006*, pp. 404–411, 2006.
15) H.Narimatsu, M.Nakano, and K.Funakoshi. A classifier-based approach to supporting the augmentation of the question-answer database for spoken dialogue systems. In G.G. Lee, J.Mariani, W.Minker, and S.Nakamura eds., *Spoken Dialogue Systems for Ambient Environments: Seond International Workshop, IWSDS 2010, Gotemba, Shizuoka, Japan, October 1-2, 2010*, pp. 182–187. Springer, 2010.
16) I.O'Neill, P.Hanna, X.Liu, and M.McTear. Cross domain dialogue modelling: an object-based approach. In *Proc. Interspeech-2004 (ICSLP)*, pp. 205–208, 2004.
17) B. Pakucs. Towards dynamic multi-domain dialogue processing. In *Proc. Eurospeech-2003*, pp. 741–744, 2003.
18) E.-P. Salonen, M.Hartikainen, M.Turunen, J.Hakulinen, and J.A. Funk. Flexible dialogue management using distributed and dynamic dialogue control. In *Proc. Interspeech-2004 (ICSLP)*, pp. 197–200, 2004.
19) 佐藤, 中野, 松山, 駒谷, 船越, 奥乃博. 拡張性の高いマルチドメイン対話システムのための２段ドメイン選択法. 人工知能学会研究会資料 SIG-SLUD-60, 2010.
20) I.H. Witten and E.Frank. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann, San Francisco, 2005.