様式C-19

科学研究費補助金研究成果報告書

平成２１年６月２日現在

研究種目： 若手研究（B）
研究期間： 2007～2008
課題番号： １９７００１４２
研究課題名（和文） 実ユーザの多様な発話表現に頑健な
マルチドメイン音声対話システムの研究
研究課題名（英文） Study on multi-domain spoken dialogue systems robust against various utterance expressions of real users

研究代表者
駒谷 和範（KOMATANI KAZUNORI）
京都大学・情報学研究科・助教
研究者番号： 40362579

研究成果の概要（200字程度）：
本研究では，実ユーザの多様な言語表現に起因する音声認識誤りに対しても誤動作せず，適切な応答を行えるマルチドメイン音声対話システムの実現を目指して研究を行った．具体的には，まずシステムにおける対話履歴をモデル化し，それを利用した音声理解を実現した．次に発話検証技術を想定外発話の検出に用いた．さらにLatent Semantic Model (LSM) を導入することで，想定外発話からも頑健にトピック推定を行った．

交付額

（金額単位：円）

	直接経費	間接経費	合　計
２００７年度	2,000,000	0	2,000,000
２００８年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総　計	3,300,000	390,000	3,690,000

研究分野： 総合領域
科研費の分科・細目： 情報学・知能情報学
キーワード： 人工知能，音声対話システム，音声認識，トピック推定，Latent Semantic Mapping，対話履歴，マルチドメイン

１．研究開始当初の背景

(1) これまでに国内外の様々な研究機関で多くの音声対話システムが試作されてきた．国内では，東芝で開発されたTOSBURG（「不特定話者音声対話システム TOSBURG の開発」，竹林洋一ら他９名：日本音響学会春季講演論文集 pp.135-136 (1992)）をはじめ，NTTや京大，豊橋技科大，東大，奈良先端大などのシステムが挙げられる．同じ頃欧米では，DARPAなどの大規模プロジェクトにおいて音声対話システムの競争的開発が行われ，データの蓄積やモデルの開発などにより音声認識精度の向上が得られ，それらに基づくシステムも多く作られた（“DARPA ATIS Test Results”，Pallett, D. 他３名：Proc. Speech and Natural Language Workshop, pp.114-121 (1990)，など）．

(2) 前述した従来の音声対話システムは，基本的に対象が単一のドメイン（飛行機の予約，天気予報案内，観光案内，など）であった．単一ドメインシステムが数多く開発されている現状を踏まえ，これらを統合し，より広

い範囲を扱えるシステムを構築する時期に来ている．しかしその方法論は明らかではない．そこで，現存するシステムを再利用し，一つのインタフェースで複数のドメイン（＝既存サブシステム）を扱うシステムを構築する手法を開発する．

(3)複数のドメインを扱う音声対話システム（以後**マルチドメイン音声対話システム**）は，従来の単一ドメインシステムを複数並列に配置して利用する場合と比較して，ユーザがあらかじめシステム（ドメイン）を選ぶ必要がない点で優位である．なぜならユーザにとっては，自分の要求がどのドメインで受理できるかがわからない場合や，要求自体が複数のドメインにまたがる場合も扱えるためである．（これは，我々が日常カスタマーサポートに電話をする際に，電話番号が細かく分けられていると，どれにかけてよいかわからず困る，という状況に相当する．）さらにこの状況はドメイン数が増加するほど顕著となる．

一方マルチドメインシステムでは，ユーザの発話がどのドメイン（サブシステム）に対するものかを，システムが適切に判別する必要がある．特に，対象ドメインの数が増えることで音声認識の対象語彙も増大し，音声認識精度は低下する．またドメインが増加すると抽出すべき意図の種類も増加し，曖昧性が生じる．したがって，マルチドメインシステムでは，いかに音声認識誤りや意味解釈の曖昧性が存在する中で，頑健に対話を進められるかが重要な課題となる．

(4) 我々は，京都市バス運行情報案内システムを構築し公開してきた（075-326-3116)．本システムでは，ユーザが利用したいバス停名や系統番号を発話すると，それを音声認識・解釈し，該当するバスが近づいているかどうかが出力される．本システムは一般に公開されていた．

この公開システムから得られたデータは，多人数かつ多様なユーザによるものであり，研究室内で行う実験では得られない，一般ユーザの様々なふるまいが含まれている．具体的には，ユーザの発話内容や発話表現は多岐にわたり，これらへの対処なくして現実のユーザに対して頑健に動作するシステムは構築し得ない．したがって，実ユーザを対象とする場合では，ユーザ発話の多様性を考慮したシステム設計が不可欠である．

２．研究の目的

(1) 本研究では，ユーザの多様な発話内容に対処する一環としてマルチドメインシステムに取り組み，かつそこで**多様な言語表現**に対しても誤動作しない対話管理法の確立を目指す．具体的には，5 つ（以上）のドメインから成るマルチドメイン音声対話システムにおいて，ユーザの多様な言語表現に対して頑健な音声理解・対話管理手法を開発する．ユーザの発話がシステムにとって解釈不能であった場合でも，音声認識誤りによる誤解の発生を防いだうえで，トピックの推定により次発話に対する音声認識用言語モデルを適応させたり，発話内容に応じてシステムの受理可能な範囲内へとユーザ発話を誘導したりする．これにより，事前教示を全く与えないユーザ（≒実ユーザ）に対するタスク達成率の向上を図り，被験者を用いた評価実験により提案手法の有効性を示す．

(2) 本研究の学術的な特色・独創的な点として以下の2点が挙げられる．

①**実ユーザの発話の多様性に頑健な言語理解・対話管理**： 従来行われているマルチドメイン音声対話システムの研究として，Bor-shen Lin 他２名による“A distributed agent architecture for intelligent multi-domain spoken dialogue systems”, IEICE Trans. on Information and Systems, E84-D(9) pp.1217-1230 (2001) や Ian O'Neil 他３名による“Cross domain dialogue modeling: An object-based approach”, Proc. of ICSLP pp.205-208 (2004) などがある．しかしこれらの研究は，マルチドメインシステムのアーキテクチャに焦点を置いており，多様なユーザ発話による音声認識誤りについては考慮していない．

一般ユーザの発話にも頑健に動作するシステムを構築するには，音声認識誤りのない発話を前提とするのは不適切である．実ユーザの発話は多岐にわたり，研究室で実験を行う際の統制されたユーザの発話とは大きく異なる．

本研究では，実ユーザによる，整っていない自発的な発話に対しても，適切に言語理解・対話管理を行う方法を開発する．音声認識は入力音声に対してシステムの持つ語彙の中で最尤の単語列を出力するプロセスであるため，音声認識結果の正誤を判定するのも自明ではない．そのような中で，まずユーザの発話自体がシステムにとって正しく認識・解釈可能な範囲内であるかを推定する．さらにそのユーザ発話の内容がどのドメインに近いかを推定し，システムの解釈可能な範囲内へユーザ発話を誘導するという戦略を実現する．

②**実際の使用状況を想定した条件での被験者実験を通したシステム評価**： 本研究は，実ユーザの多様な発話に対する頑健さを志向しているため，評価実験には一般の被験者

を用いる．従来の被験者実験では，受理される発話パターンや内容を事前にユーザに教示することが多いが，この場合ユーザの発話表現が教示により影響されてしまう．本研究では，一般ユーザが全くの教示なしにシステムを使うような状況を想定し，実験条件を設定して評価実験を行う．

３．研究の方法

本研究の目的は，複数のドメインに対する音声対話システムにおいて，ユーザの発話の多様性に対しても頑健な言語理解・対話管理の実現である．このために，(1) 機械学習による対話履歴を利用した言語理解，(2) 発話検証技術を利用した，ユーザ発話のシステム適合性の判別，(3) LSM(Latent Semantic Mapping)に基づくトピック推定，などの技術を音声対話システムに新たに導入し，音声対話システムにおける新たな言語理解や対話管理の手法を開発する．

これらの技術により得られる情報は，音声対話システムにおける言語理解部や対話管理部において利用され，ユーザに適切にフィードバックされる．このような対話戦略をマルチドメイン音声対話システムにおいて実現したうえで，これを音声対話システムにより収集した実データにおいて検証する．

具体的には下記の各点について検討・実装を進める．

【平成19年度】

(1) マルチドメイン音声対話システムにおけるデータ収集

まずマルチドメイン音声対話システムにおいてデータ収集を行う．より実データに近い発話を得るために，音声対話システム使用経験のない被験者に対して，ほぼ教示を行わずにシステムを使用してもらうことでデータを収集する．この際の被験者に対して謝金を支給する．また得られた音声データの書き起こしにはアルバイトを活用する．

(2) 対話履歴を利用した音声認識結果の頑健な解釈

従来から用いられる，音響情報や言語情報などボトムアップ情報に基づく音声認識の信頼度に加え，対話の文脈的制約をトップダウン情報としてモデル化し，音声認識結果の解釈に導入する．これにより，文脈から外れた音声認識誤りの棄却や，文脈情報を用いた発話解釈の曖昧性の解消が可能となる．

(3)発話検証技術の利用法の開発

ユーザ発話自体が音声対話システムにとって受理可能であるかどうかを判別する．もしユーザ発話がシステムの受理可能な範囲外である場合，音声認識結果は必ず誤りを含み，かつその場合ユーザに単純に再発話を要求しても同じ誤りが繰り返されるため意味がない．そこで，発話検証技術を適用することで，ユーザの発話が文法に沿っているかを推定する方法を開発する．

【平成20年度】

(4)LSM(Latent Semantic Mapping)に基づくユーザ発話のトピック推定

音声対話システムでは，タスク遂行に必要な単語やフレーズを確実に認識，解釈するために，音声認識部の語彙サイズを絞ることが多い．このため，ユーザの発話がシステムの想定範囲外である場合の対処が必須である．平成19年度に開発する発話検証技術を用いてこれを範囲外と判定し棄却するだけでなく，その後のユーザ発話の適切な誘導や，次発話に対する音声認識語彙を絞り込めるのが望ましい．本研究では，より広範囲な言語モデルを用いて並行して音声認識を行い，その結果に対してLSMを適用することで，その発話のトピックを推定する．LSMは情報検索で用いられる，次元圧縮を用いた文書分類手法である（“Latent semantic mapping”, Bellegarda, J.R., Signal Processing Magazine IEEE, Vol.22 Issue 5, pp.70-80 (2005)）．

(5)各特徴の統合による対話戦略の開発，評価実験

上記の各技術により得られる情報に基づいて，統合的な対話戦略を開発し，マルチドメイン音声対話システムに組み込む．これにより，ユーザの発話表現が多様であっても，システムは適切に対処できるようになる．具体的には想定外発話に対しても，このシステムに対して被験者を用いて評価を行い，開発した方式の有効性を示す．この際，（研究室のメンバーではない）一般の被験者の発話を実験に用いることで，より実ユーザに近い発話に対して性能を評価する．

４．研究成果

(1) 対話履歴を利用した音声認識結果の頑健な解釈

音響情報や言語情報などボトムアップ情報に基づく音声認識の信頼度に加え，対話の文脈的制約をトップダウン情報としてモデル化し，音声認識結果の解釈に導入した．具体的には，マルチドメインシステムにおけるドメインの遷移を，(i)ひとつ前の応答を行ったドメイン，(ii)音声認識結果を受理できる最尤のドメイン，(iii)上記以外のドメイン，

と抽象化し，対話から得られる様々な特徴を用いて決定木学習により判別する手法を開発した．本手法を実装して評価実験を行い，ドメイン選択誤りが 18.2%削減されることを確認した．これにより本手法の有効性を示した．

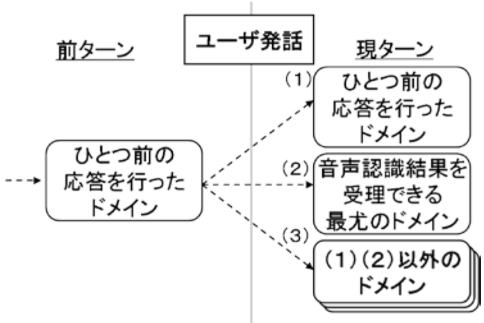

図 1：対話履歴を考慮したドメイン選択

(2) 発話検証技術の利用法の開発
ユーザ発話自体が音声対話システムにとって受理可能であるかどうかを判別する手法を開発した．もしユーザ発話がシステムの受理可能な範囲外である場合，音声認識は必ず誤りを含み，かつその場合ユーザに単純に再発話を要求しても意味がない．そこで，発話検証技術を適用し，複数のデコーダによる音響尤度を比較することで，ユーザの発話が文法に沿っているかを推定する方法を開発した．評価実験をユーザの自由な発話に対して行い，有効性が示されたが，短い発話に対しては文法内外の判定を誤る場合が多く，この点に関する改善の必要性も明らかになった．

(3) LSM(Latent Semantic Mapping)を用いたユーザ発話のトピック推定
LSM を用い，システム想定外発話に対しても頑健にトピックを推定する手法を開発した．ユーザがシステムの受理できない発話をした場合，システムは誤動作しないだけでなく，その発話のトピックを推定することができれば，その後の対話管理に有用である．そこで Web から大量かつ半自動的にテキストを収集し，さらにそこに含まれるノイズの影響を取り除くために LSM を導入した．

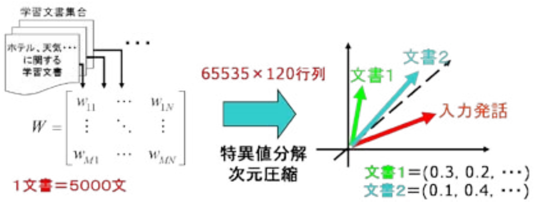

図 2： LSM による各トピックの表現

(4) 対話履歴とトピック推定結果を統合したドメイン選択の実現

上記のトピック推定結果と対話履歴を決定木学習により統合し，ドメイン選択精度の向上を試みた．これによるドメイン選択精度の向上を評価実験により示した．またこの際に対話履歴とトピック推定結果が相補的な情報を持つことも明らかになった．

(5) 文法検証結果を用いたヘルプ候補のランキング
ユーザの発話が想定外である場合に，システム側からユーザを誘導するヘルプメッセージを生成する．この際にユーザ発話の文法検証結果と当該ユーザの発話履歴を統合してヘルプ候補をランキングする手法を開発した．この際の正解ラベルは人手でも一意に決めることは難しいため，複数の正解ラベルを付与し，これに基づきブースティングアルゴリズムを用いることでランキングを行った．合計 30 名の一般ユーザから収集した発話データを用いた評価実験により，本手法により正しいヘルプ候補をより上位にランキングできることを示した．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計４件）

① 池田 智志，駒谷 和範，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．情報処理学会論文誌, Vol.50, No.2, pp. 488--500, 2009．査読有
② Kazunori Komatani, Satoshi Ikeda, Tetsuya Ogata, Hiroshi G. Okuno: Managing Out-of-Grammar Utterances by Topic Estimation with Domain Extensibility in Multi-Domain Spoken Dialogue Systems. Speech Communication, Vol.50, Issue 10, pp.863--870, 2008．査読有
③ 福林 雄一朗，駒谷 和範，中野 幹生，船越 孝太郎，辻野 広司，尾形 哲也，奥乃 博：音声対話システムにおけるラピッドプロトタイピングを指向した言語理解．情報処理学会論文誌，Vol.49，No.8，pp.2762--2772，2008．査読有
④ 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．査読有

〔学会発表〕（計１４件）

① 池田 智志, 駒谷 和範, 高橋 徹, 尾形 哲也, 奥乃 博： 音声対話システムにおける想定外発話の文法検証を用いた対話行為推定に基づくヘルプ生成, 情報処理学会第 71 回全国大会, 4Q-4, Mar. 2009. 査読無

② Satoshi Ikeda, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Extensibility Verification of Robust Domain Selection against Out-of-Grammar Utterances in Multi-Domain Spoken Dialogue System. Interspeech-2008, pp.487--490, 2008. 査読有

③ Satoshi Ikeda, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Integrating Topic Estimation and Dialogue History for Domain Selection in Multi-Domain Spoken Dialogue Systems. The 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligence Systems (IEA/AIE-2008), LNAI5027, pp.294--304, 2008. 査読有

④ 池田 智志，駒谷 和範，尾形 哲也，奥乃 博： トピック推定と対話履歴の統合によるドメイン選択を行うマルチドメイン音声対話システム．情報処理学会第 70 回全国大会，3U-3，2008．査読無

⑤ 福林 雄一朗，駒谷 和範，尾形 哲也，奥乃 博： 音声対話システムにおける WFST に基づく文法検証を利用した動的ヘルプ生成． 情報処理学会第 70 回全国大会，3U-7，2008． 査読無

⑥ 福林 雄一朗，駒谷 和範，尾形 哲也，奥乃 博： 音声対話システムにおける動的ヘルプ生成を指向した WFST に基づく文法検証によるユーザ知識推定．人工知能学会研究会資料， pp.45--50， 2008． 査読無

⑦ 池田 智志，駒谷 和範，尾形 哲也，奥乃 博： 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． 情報処理学会研究会資料，pp.277--282，2007． 査読無

⑧ 池田 智志，駒谷 和範，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想定外発話への対処のための Web を用いたシステム知識の拡張． 情報技術レターズ， Vol. 6, pp.131-132，2007．査読有

⑨ 福林 雄一朗，駒谷 和範，中野 幹生，船越 孝太郎，辻野 広司，尾形 哲也，奥乃 博： 音声認識結果とコンセプトへの重みづけによる WFST に基づく音声言語理解の高精度化．情報技術レターズ, Vol. 6, pp.133-134，2007．査読有

⑩ 池田 智志，駒谷 和範，尾形 哲也，奥乃 博： 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．人工知能学会研究会資料， pp.83--88， 2007． 査読無

⑪ 福林 雄一朗，駒谷 和範，中野 幹生，船越 孝太郎，辻野 広司，尾形 哲也，奥乃 博： 音声認識結果とコンセプトへの重みづけによる WFST に基づく音声言語理解の高精度化． 情報処理学会研究会資料， pp.43--48， 2007．査読無

⑫ Yuichiro Fukubayashi, Kazunori Komatani, Mikio Nakano, Kotaro Funakoshi, Hiroshi Tsujino, Tetsuya Ogata, Hiroshi G. Okuno: Rapid Prototyping of Robust Language Understanding Modules for Spoken Dialogue Systems. The Third International Joint Conference on Natural Language Processing (IJCNLP2008), pp.210--216, 2008. 査読有

⑬ Kazunori Komatani, Yuichiro Fukubayashi, Tetsuya Ogata, Hiroshi G. Okuno: Introducing Utterance Verification in Spoken Dialogue System to Improve Dynamic Help Generation for Novice Users. 8th SIGdial Workshop on Discourse and Dialogue, pp.202--205, 2007. 査読有

⑭ Satoshi Ikeda, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Topic Estimation with Domain Extensibility for Guiding User's Out-of-Grammar Utterances in Multi-Domain Spoken Dialogue Systems. Interspeech-2007, pp.2561--2564, 2007. 査読有

６．研究組織

(1)研究代表者

駒谷　和範（KOMATANI KAZUNORI）
京都大学・情報学研究科・助教
研究者番号： 40362579

(2)研究分担者

該当なし

(3)連携研究者

該当なし