

機関番号：13901

研究種目：若手研究(B)

研究期間：2009～2010

課題番号：21700164

研究課題名(和文) 音声対話システムにおけるユーザ行動の時系列変化を捉えたユーザモデリング

研究課題名(英文) User Modeling on Temporal Changes in User Behaviors for Spoken Dialogue Systems

研究代表者：

駒谷 和範 (KOMATANI KAZUNORI)

名古屋大学・工学研究科・准教授

研究者番号：40362579

研究成果の概要(和文)：本研究では、音声対話システムにおけるユーザの発話行動の経時的変化に着目し、ユーザのシステムへの習熟過程のモデル化を目指して研究を行った。これを通じてユーザの発話行動(音声認識率、バージイン率など)とその変化に関する知見を蓄積し、ユーザに適応したインタラクションのための基礎技術とする。この習熟過程のモデルは、各ユーザに適応した応答生成や音声認識結果の取捨選択に利用可能である。本研究ではその有効性をバージイン発話の取捨選択精度の向上により示した。

研究成果の概要(英文)：We quantitatively modeled users' habituation to a spoken dialogue system by focusing on temporal changes in user behaviors measured by automatic speech recognition (ASR) accuracy and barge-in rates. This model on changing user behaviors can be used as a fundamental finding for realizing user-adaptive interactions. The model is applicable to spoken dialogue systems that generate adaptive responses and select correct ASR results from those that may contain errors. We showed its effectiveness by the improvement of selection accuracy for barge-in utterances.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,800,000	540,000	2,340,000
2010年度	1,500,000	450,000	1,950,000
総計	3,300,000	990,000	4,290,000

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：音声対話システム、ユーザモデリング、経時的変化、音声認識、習熟度、バージイン、信頼度、発話タイミング

1. 研究開始当初の背景

(1) 現状の音声対話システムに欠けている要素のひとつに、各ユーザに適応したインタラクションを行う能力が挙げられる。音声メディアを用いたコミュニケーションでは、一度に伝達可能な情報量が限られるため、相手の必要とする情報のみを選択的に伝えることが、画像などの他メディアを用いる場合と比べて重要となる。とりわけ一般ユーザにシステムを公開する場合には、システムに対する知識やそれに伴う発話行動がユーザごとに大きく異なるため、適応

的な対話戦略が不可欠である。

これまでも国内・国外ともに、多くの音声対話システムが構築されてきた。我々が開発した京都市バス運行情報案内システム[1]や、カーネギーメロン大学での Let's Go! バスシステム[2]、翠・河原らによる音声対話型京都案内システム[3]、奈良先端大のたけまるくん[4]などが一般ユーザに公開され、発話が収集されてきた。

これら現在の音声対話システムは、あらゆるユーザに対して平均的に動作するように設計されている。また、Partially

Observable Markov Decision Process (POMDP) に基づく確率的な対話管理[5]が盛んに研究されつつあるが、これを用いた研究の多くも全てのユーザに対して平均的に最適な対話戦略を学習するものである。このように個々のユーザに適応する技術は必ずしも研究が進んでいない状況であった。

[1] K. Komatani, et al., “User modeling in spoken dialogue systems to generate flexible guidance,” User Modeling and User-Adapted Interaction, vol. 15, no. 1, pp. 169--183, 2005

[2] A. Raux, et al., “Doing research on a deployed spoken dialogue system: One year of Let’s Go! experience,” Proc. INTERSPEECH, pp. 65-68, 2006.

[3] 翠輝久, 他: 質問応答・情報推薦機能を備えた音声による情報案内システム. 情報処理学会論文誌, Vol. 48, No. 12, pp. 3602--3611, 2007.

[4] 鹿野清宏, 他: 音声情報案内システム「たけまるくん」および「キタちゃん」の開発, 情報処理学会研究報告, 2006-SLP-63-(7), pp. 33--38, 2006.

[5] Jason D. Williams and Steve Young: Partially Observable Markov Decision Processes for Spoken Dialog Systems. Computer Speech and Language 21(2): 393-422, 2006.

(2) 研究代表者を含むグループでは、2002年5月から2007年3月まで、京都市バス運行情報案内システムを一般市民に公開し、多様なユーザの発話を含む対話データを収集してきた。その分析を進めるなかで、**様々なユーザに適応した処理は必須**であると強く考えるに至った。

例えば、まず**初心者ユーザにはシステムの使い方に関する説明が不可欠**である。音声インタフェースでは、Webなどの視覚的なインタフェースとは異なり、システムが受理できる言語表現やシステムの状態を逐一伝えられない。このためユーザがシステムの受理可能な表現を知らない場合には、音声認識誤りが繰り返され、対話が破綻してしまう。その一方、**慣れたユーザは早く情報を得ようとするため、既知の説明を長々と行うと素早いタスク遂行が不可能**になる。

さらにユーザの発話行動の多様さは各個人間の差だけにとどまらず、**同じ個人でも慣れにより発話行動は変化する**。単純に各ユーザへの適応を考えるだけでなく、各ユーザの慣れも考慮した経時的なモデル化が必須であった。

2. 研究の目的

本研究課題では、京都市バス運行情報案内システムで収集したデータをもとに、音声対話システムに対する**実ユーザの習熟過程のモデル化**を行う。これを通じてユーザの発話行動（音声認識率、バージン率など）とその変化に関する知見を蓄積し、**ユーザに適応したインタラクションを実現する際の基礎技術**を確立する。

我々は以前にも音声対話システムにおけるユーザモデリングの研究を行ってきた[1]。その研究では、各対話で得られる特徴のみを用いて、システムの応答生成をユーザに適応させた。本研究課題は[1]を以下の2点において展開するものである。

(1) ユーザ個人ごとの慣れによる発話行動の変化の分析

ユーザが繰り返しシステムを使う状況において、一回の対話内だけでなく、各ユーザの発話行動を経時的に分析することで、ユーザの発話行動の変化をモデル化し、予測に用いる。この分析に、大量の実データを用いる。本研究課題で使用した京都市バス運行情報案内システムのデータは、一般市民から継続的に収集された。この中にはシステムを繰り返し使用したユーザが少なからず存在し、ユーザの習熟過程のモデル化に利用できる。カーネギーメロン大学のLet’s Go!バスシステム[2]も一般市民に公開されているが、ユーザごとの分析は行われていない。音声対話システムでの大規模データに基づく経時的なユーザ行動の分

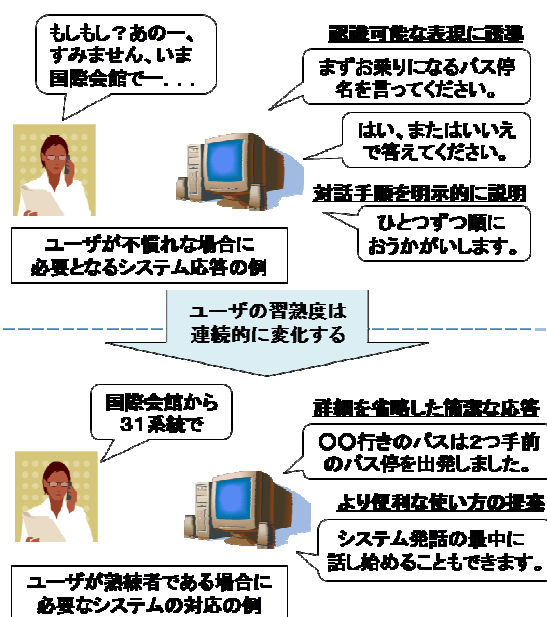


図1: 適応的な応答の例

析・モデル化は、世界的にも類を見ない。

(2) 音声認識結果の取捨選択など他レベルの技術との統合

ユーザの習熟度のモデルを、応答生成の適応だけでなく、音声認識結果の取捨選択にも適用する。これにより、構築したモデルの有効性を、従来の音声理解の要素技術の精度向上として示すことができる。また従来のユーザモデリングの研究は、主に発話内容からユーザを分類するものが多い[6]。本研究課題では、音声認識率やタスク達成率、バージン率など、発話行動自体を利用するため、発話内容（音声認識結果）を利用するよりも音声認識誤りの影響を受けにくい。このモデルから得られる情報を、音声認識結果の取捨選択など言語理解レベルの処理に活用することで、音声認識結果に誤りが含まれる場合でも、適切な対話を続けられるようになる。

[6] R. Kass and T. Finin: Modeling the user in natural language systems. Computational Linguistics, vol.14, No.3, pp.5-22, 1988.

3. 研究の方法

本研究では、京都大学で収集された、京都市バス運行情報案内システムのデータに対して分析を行った。これは2002年5月から2007年3月までの59ヵ月分のデータである。このデータでは、大部分のコールで発信者番号が記録されているため、この番号を各ユーザとみなして発話行動を分析した。まず平成21年度に、このデータに基づいてユーザのふるまいを分析し、その経時的な変化を定量的に示した。続いて平成22年度には、推定音声認識率の設計とそれを用いた音声認識結果の取捨選択について検討し、検証を進めた。

【平成21年度】

(1) 実ユーザの発話行動の時系列的な分析

音声認識率、タスク達成率、バージン率などを各個人ごとに算出し、その時系列変化を分析する。これにより、各ユーザの発話行動の時系列的な変化の有無やその度合、大きさなどを定量的に調べ、ユーザが長期間システムを使用した際の発話行動の変化をモデル化する。

(2) 各時系列データ間の依存関係の分析とモデル化

ユーザの発話行動の各指標間の関係を分析する。例えば、音声認識率とタスク達成率や、音声認識率とバージン率の関係などを調査し、変化の順序などに傾向がないかを分析する。特にバージン発話（システム発話を遮って行われたユーザ発話）の音声認識率が低いことが分かっているため、各ユーザのバー

ジン率を用いた音声認識誤りの予測等、その関係に基づく発話理解への発展を検討する。

【平成22年度】

(3) ユーザごとの音声認識率の推定

各ユーザの音声認識率は、それぞれの習熟度合を示す尺度として有用であることが、前年度の検討によりわかった。しかしこれを求めるには、事後的に人手による書き起こしを行う必要である。つまり全てのユーザ発話に対して書き起こしを行う労力が必要であり、その結果オンラインでシステムがユーザに適応することができないことになる。これを克服するため、対話中に生じる特徴を使って、ユーザの音声認識率をオンラインで計算する方法を検討する。具体的には、音声対話システムにおいて頻発する明示的確認に対する応答（つまり「はい」や「いいえ」）に着目し、これを活用する方法を検討する。

(4) 音声認識結果の取捨選択への適用

従来、音声認識結果は音声認識の信頼度に対するしきい値処理で取捨選択される。このしきい値をユーザごとに適応させることで選択精度の向上を図る。さらに音声認識結果の取捨選択精度の向上を目指し、バージン率を信頼度計算の特徴に加えることで、ユーザの発話行動の特徴を考慮した新たな信頼度計算法を検討する。さらに、バージンの有無だけでなく、割り込んだタイミングを考慮して信頼度を計算することも検討する。

4. 研究成果

【平成21年度】

(1) 実ユーザの発話行動の経時的な分析

音声認識率、タスク達成率、バージン率などを各個人ごとに算出し、その時系列変化を分析した。ここでは記録されていた電話番号をユーザとみなした。これは電話番号の多くが携帯電話のものであったことから妥当な仮定である。これにより、各ユーザの発話行動の時系列的な変化の有無やその度合、大きさなどを定量的に検証した。これらはユーザのシステムに対する習熟度合に対応するとみなすことができ、これに応じた応答生成や発話の取捨選択などに活用できる。

(2) 各時系列データ間の依存関係の分析とモデル化

上記で示したユーザの発話行動の各指標間の関係を分析した。この結果、音声認識率とタスク達成率はほぼ同時に向上することや、音声認識率とバージン率では後者が遅れて向上することを定量的に確認した。この

習熟の順序は、Walkerら(1997)による音声対話システムの評価指標(PARADISE)における、

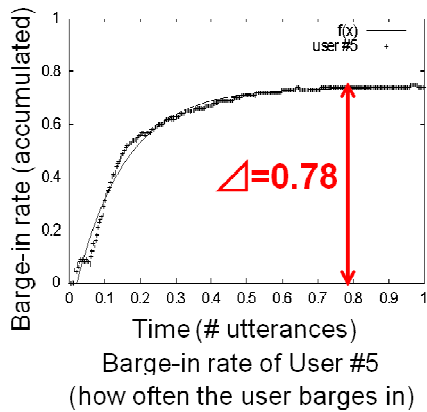


図2: あるユーザのバージン率の変化

ユーザの主観的評価に対して影響の強い項目の順番に対応する。この結果は、ユーザがシステム評価時において重要視する要素が、実際に一般ユーザの習熟過程における順序と対応するということがデータから実験的に示されており、興味深い結果である。

【平成22年度】

(3) 推定音声認識率の設計

音声対話システム使用中の各ユーザの音声認識率を、人手による書き起こし無しで推定する手法を検討・開発した。通常、音声認識率は発話を人手で書き起こした結果を用いて計算するため、システム動作中にこのような情報を使用することはできない。本研究では各ユーザの音声認識精度を、肯定否定発話に着目してオンラインで推定する手法を考案した。具体的には、明示的確認に対する応答の音声認識率が高いことに基づいて、これを正しいと仮定し、また「はい」などの肯定応答に対応するユーザ発話の音声認識結果も正しかったと仮定する。これらの仮定に基づき当該ユーザの音声認識率を計算し、これを推定音声認識率とした。

図3に、この推定音声認識率と、書き起こしに基づく音声認識率の相関を示す。相関係数は0.806と比較的高く、推定した音声認識率が実際の音声認識率と高い相関をもっていることを確認した。さらにこれをバージン発話の取捨選択に用いた結果、取捨選択精度の向上が得られ、ユーザプロファイルとして利用可能であることが示された。また従来用いられる音声認識信頼度とともに用いた場合でも精度が向上した。

(4) 音声認識結果の取捨選択への適用

当該ユーザのバージン率や推定音声認識率をロジスティック回帰により統合して音声認識信頼度として扱い、これを用いてバージン発話の取捨選択を行った。バージン

率が、同一ユーザでも変化することを考慮したモデル化により、予測精度が向上した。またこれらバージン率や推定音声認識率は、従来から用いられている音声認識信頼度とともに用いた場合でも有効であり、これら対話レベルから得られる情報として有効であることも示された。

さらに、単純にバージンの有無を尺度として用いるだけでなく、ユーザがシステム発

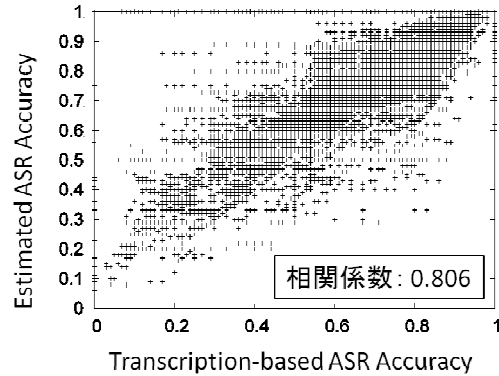


図3: 推定音声認識率と実際の音声認識率との相関

話のどのタイミングで割り込んだかを特徴として加えた。この際に、ユーザ発話の割り込みのパターンを分類し、発話の衝突や、発話区間検出誤りによる一発話の分割が起きていることを明らかにした。またバージン発話の取捨選択精度により評価し、バージンタイミングが発話の取捨選択精度の向上に有効な情報であることを示した。

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)

- ① [雑誌論文] (計3件)
- ② Mikio Nakano, Yuji Hasegawa, Kotaro Funakoshi, Johane Takeuchi, Toyotaka Torii, Kazuhiro Nakadai, Naoyuki Kanda, Kazunori Komatani, Hiroshi G Okuno, Hiroshi Tsujino: A multi-expert model for dialogue and behavior control of conversational robots and agents. Knowledge-Based Systems, 査読有, Vol.24, Issue 2, 2011, pp.248--256.
- ③ Kazunori Komatani, Yuichiro Fukubayashi, Satoshi Ikeda, Tetsuya Ogata, Hiroshi G. Okuno: Selecting Help Messages by using Robust Grammar Verification for Handling Out-of-Grammar Utterances in Spoken Dialogue Systems. IEICE Transactions

on Information and Systems, 査読有, Vol. E93-D, No. 12, 2010, pp. 3359-3367.

- ④ 勝丸真樹, 中野幹生, 駒谷和範, 船越孝太郎, 辻野広司, 尾形哲也, 奥乃博: 複数の言語モデルと言語理解モデルによる音声理解の高精度化. 電子情報通信学会論文誌, 査読有, Vol. J93-D, No. 6, 2010, pp. 879-888.

[学会発表] (計7件)

- ① 中島 大一, 駒谷 和範, 佐藤 理史: 音声対話システムにおけるバグイン発話の分類とそれに基づくエラー検出. 言語処理学会第17回年次大会講演論文集, 2011年3月8日, 豊橋技術科学大学 (愛知県).
- ② 駒谷 和範, 松山 匡子, 奥乃 博: 発語行為レベルの情報を活用した音声対話システムの研究. 人工知能学会研究会資料, SIG-SLUD-B002-12, pp. 53--58, 2010年10月28日, 早稲田大学 (東京都).
- ③ Kazunori Komatani, Hiroshi G. Okuno: Online Error Detection of Barge-In Utterances by Using Individual Users' Utterance Histories in Spoken Dialogue System. 11th Annual SIGDIAL Meeting on Discourse and Dialogue, 査読有, pp. 289--296, Sep. 25, 2010, 東京大学 (東京都).
- ④ 駒谷 和範, 奥乃 博: 音声対話システムにおける各ユーザの利用履歴を活用したバグイン発話のエラー検出. 言語処理学会第16回年次大会講演論文集, 査読無, pp. 218--221, 2010年3月9日, 東京大学 (東京都).
- ⑤ 駒谷 和範, Alexander I. Rudnicky: 音声対話システムにおける暗黙的な教師信号に基づく音声認識率の推定とそれを用いたエラー予測. 情報処理学会研究報告, 査読無, 2009-SLP-78-3, 2009年10月23日, 早稲田大学 (東京都).
- ⑥ Kazunori Komatani, Alexander I. Rudnicky: Predicting Barge-in Utterance Errors by using Implicitly-Supervised ASR Accuracy and Barge-in Rate per User. ACL-IJCNLP 09, Short Papers, pp. 89--92, 査読有, Aug. 4, 2009, Singapore (Singapore).
- ⑦ Kazunori Komatani, Tatsuya Kawahara, Hiroshi G. Okuno: A Model of Temporally Changing User Behaviors in a Deployed Spoken Dialogue System. International Conference on User Modeling, Adaptation, and Personalization (UMAP09), LNCS 5535, pp. 409--414, 査読有, Jun. 23, 2009, Trento (Italy).

[その他]

ホームページ等

<http://sslslab.nuee.nagoya-u.ac.jp/~komatani/>

6. 研究組織

(1) 研究代表者

駒谷 和範 (KOMATANI KAZUNORI)

名古屋大学・大学院工学研究科・准教授

研究者番号: 40362579

(2) 研究分担者 なし

(3) 連携研究者 なし