

インタラクティブエージェント用 ユーザモデル構築のための対話実験

A Wizard of Oz Experiment for Constructing A User Model of Interactive Agents

松本 泰明^{*1*2} 麻生 英樹^{*2} 原 功^{*2} 柿倉正義^{*1}
Yasuaki MATSUMOTO Hideki ASOH Isao HARA Masayoshi KAKIKURA

^{*1}東京電機大学 ^{*2}産業技術総合研究所
Tokyo Denki University AIST

User models are a key element of user adaptation of dialog systems. In the previous work, we have proposed a probabilistic user model using a dynamic Bayesian network for estimating users' intention. This time, we extend the model to incorporate users' profile. In order to choose appropriate features for estimating the profile and to estimate the probability distribution of the features and profile, dialog data is necessary. Here, we introduce a Wizard of Oz type system for gathering dialog data, and describe an experiment using the system. Preliminary results show that a part of user's profile can be estimated from the dialog data.

1. はじめに

ロボット技術やソフトウェアエージェント技術の発達に伴い、音声やジェスチャなどのユーザに親しみやすいモダリティを用いて、インタラクティブに各種サービスを行うことが可能な対話システムや対話ロボットの研究開発が盛んになっている [加藤 95, 鹿野 02, 松阪 02]。しかし、現在開発されている対話システムは、想定したシナリオから対話がはずれないように対話の主導権を常にシステムがとるものや、逆に、ユーザからの質問に対する応答を反射的に出力するだけというシステムが多い。そのため、ユーザは対話中に単調なやり取りを繰り返す必要があったり、提供された情報がそのユーザにとって有用でない場合があったりするなど、まだまだ使い勝手がよいものにはなっていない。

我々は、柔軟な対話を実現するためのアプローチの一つとして、確率統計的な手法を利用したユーザモデルを構築して、ユーザに適応した対話を可能にすることをめざした研究を進めている。システムは、あらかじめ学習用に採取されたデータや、ユーザとの対話中に得られる情報から、対話中のユーザのプロファイルや状況、意図を推測し、その結果を用いて適切な対話の制御や応答の生成を行う。これにより、音声対話システムに使い慣れていないユーザに対してはタスク達成率を上げ、システムに慣れているユーザに対しては少ない対話のターン数で欲しいサービスを提供することが可能になると期待される。

[麻生 03] では、このような対話を実現するために、ダイナミックベイジアンネットワークによって記述されたユーザモデルを用いた確率推論によって、ユーザの意図を逐次推測し、対話の制御を行う方法を提案した。そこではユーザプロファイルは考慮していなかったが、ユーザモデルにダイナミックベイジアンネットワークを用いることの1つの利点は、ネットワークに必要な変数のノードを追加することによって、容易にモデルを拡張できることにあり、対話中の情報からユーザのプロファイルを推定し、それを対話に利用するための枠組みを作成することも簡単に行える。

しかしながら、人間が対話中に、どのような変数に着目して相手のプロファイルや状態を推定しているかは明らかにされていないため、ユーザモデルに具体的にどのような変数を追加す

ればよいかはわからない。また、追加したノードと他のノードとの確率的な関係も自明では無いことが多い。従って、ユーザに実際に対話を行ってもらい、データを収集して、有効な変数や確率分布を推定することが試みられている。

本発表では、そうした学習データの収集および、ユーザモデルの拡張を行うための準備として、Wizard of Oz (WOZ) 法による対話データ収集システムを作成し、予備的な対話実験とそれに基づくユーザプロファイルの推定を行った結果について述べる。以下では、まず、ユーザモデルにユーザのプロファイルを導入するための拡張について簡単に述べ、対話データ収集のために構築したシステムを紹介し、ユーザの習熟度や知識度を推定するための対話実験とその結果について述べる。

2. DBN による対話制御

2.1 ユーザ意図の推定

ベイジアンネットワークとは、複数の事象間の因果関係を確率によって表現する確率推論モデルであり、ユーザモデリング等に盛んに用いられている [Russel 95, 本村 03]。ベイジアンネットワーク上では、各事象をネットワークのノード、事象間の確率的な関係をノード間の有向リンクとして表現している。各事象には離散あるいは連続の状態値を持つ確率変数が割り当てられている。各ノード間の確率的な依存関係は、各ノードに定義されている条件付確率表 (Conditional Probability Table: CPT) によって表現される。ベイジアンネットワーク上での確率的推論は、観測されたいくつかの変数の確定値 (evidence: E) から、知りたい確率変数 X の事後確率分布 $P(X|E)$ を CPT とベイズの定理を基本とした計算で求めることで行われる。

システムがユーザに対して協調的な対話を行うためには、対話中のユーザがどのような特徴を持っており、何を達成しようとしているのかという対話相手に関する知識、つまり、ユーザモデルが必要である [加藤 95]。我々は、ユーザへの応答とユーザの意図とから、ユーザの反応が生成される過程をダイナミックベイジアンネットワーク (dynamic Bayesian network: DBN) を用いて表現したユーザモデルを提案した [麻生 03]。DBN とは、ベイジアンネットワークを時系列変数を扱えるように拡張したもので、通常のベイジアンネットワークを時間方向にコピーして展開した構造を持つネットワークである。

Fig.1 にユーザ意図を推定するためのネットワークの最も粗

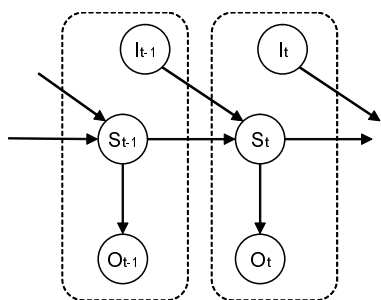


図 1: ユーザの意図推定のための確率ネットワーク

い構造を示す．隠れ変数 S_t はある時刻 t のユーザの意図（システムに対する情報の検索要求など）を表している． I_t は対話システムからユーザへの応答，つまりユーザへの入力を表す変数であり， O_t は対話システムへの入力，つまりユーザの出力を表す変数である．このようなモデルを用いれば，ユーザの意図を推定する問題は，観測可能な変数 $I_1, \dots, I_{t-1}, O_1, \dots, O_t$ から隠れ変数 S_1, \dots, S_t を推定する問題に帰着される．なお，実際にネットワークを用いる際には，各変数をさらに詳細化，具体化することになる．

2.2 ユーザ適応のためのモデルの拡張

前項のユーザモデルに対してユーザ適応を行うために，ネットワークにユーザプロファイルを推定する部分を追加し，ユーザモデルの拡張を行った．拡張されたネットワークを図 2 に示す．

今回ネットワークに新たに追加した変数は，ユーザプロファイルを表す隠れ変数 P_t と対話中に観測される変数（以下では対話パラメタと呼ぶ） C_t である．まず，ユーザの意図 S_t の親ノードにユーザプロファイル P_t を追加することで，ユーザがどのような情報を欲しているか，どのような対話を望んでいるかが，ユーザのプロファイルに依存することをモデル化した．また，ユーザプロファイル P_t に対話パラメタ C_t を子ノードとして与えることで，ユーザプロファイルによって対話パラメタの値が影響されることをモデル化している．

このモデルを用いることで，対話中に得られる対話の特徴量（対話パラメタ）をもとに，ユーザプロファイルを推測し，各ユーザに適応した対話を行うことが可能になる．実際にユーザと対話を行う場合の手順は以下ようになる．

1. 音声入力や画像入力などから観測可能な変数であるユーザの出力 O_t や対話パラメタ C_t が得られたら，直前のユーザへの入力 I_{t-1} とともに，証拠としてネットワークのノードにセットする．
2. 確率伝播計算を行い，ユーザプロファイル P_t とユーザの意図 S_t の事後確率分布を算出する．
3. 得られた算出結果から，ユーザプロファイルとユーザの意図を評価し，適切と思われるシステムの応答を生成し出力する．

このネットワークを実際の対話システムで用いるためには，ユーザプロファイル P_t や対話パラメタ C_t として，具体的にどのような変数があるかを定める必要がある．また，具体化された P_t と対話パラメタ C_t ，ユーザプロファイル P_t とユーザの意図 S_t の間の依存関係を定量化した条件付き確率表 (CPT) $P(C_t|P_t)$ ， $P(S_t|S_{t-1}, I_{t-1}, P_t)$ などを適切に設定す

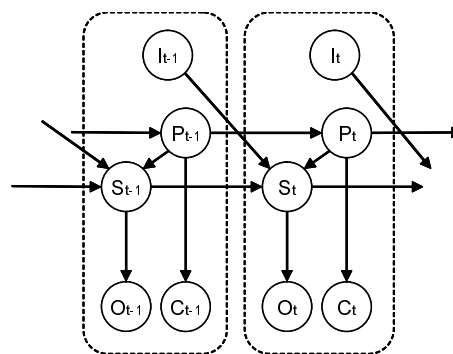


図 2: ユーザ適応のために拡張を行ったベイジアンネットワーク

必要がある．これらの選択や設定は自明ではない．そのため，実際のタスクにおいて対話データを収集して，それを学習データとして用いることにした．

3. WOZ 法による対話データ収集システム

Wizard of Oz (WOZ) 法と呼ばれる手法を使ってユーザと対話システムの実際の対話からデータを収集するためのシステムを構築した．WOZ 法とは，Wizard と呼ばれる人間のオペレータがシステムの一部，または全体を操作することでユーザとの対話を行うというものである．ユーザは，実際にシステムを相手にしていると思いながら対話を行うため，実際のシステムの運用状態に近い状況でのデータが得られる．

3.1 システム構成

今回の実験では，システムが扱うタスクとして，WWW を併用した対話的な案内タスクを想定した．ユーザからシステムへの入力は音声とし，システムからユーザへの出力は音声合成と PC ディスプレイ上のロボットシミュレータによるジェスチャ，であり，さらに，必要に応じてユーザにとって有用と思われる WWW ページを PC のディスプレイ上に表示する．

製作した対話データ収集システムの構成を Fig.3 に示す．システムは，音声認識部，音声合成部，ブラウザ制御部，ロボットシミュレータ，対話管理部，データベースの 6 つのモジュールから構成されている．各モジュールは，独立したプロセスになっており，TCP を介して互いに通信を行いながら動作をする．

音声認識部は，Julius for SAPI [Julius 04] を用いて実装しており，ユーザが発話をするとき発話内容をテキストと簡単な意味情報に変換する．音声合成部は，Microsoft 社製の音声合成エンジンを使用した．ブラウザ制御部は，ユーザに対して Web ページを表示するためのモジュールで，Windows 上の Web ブラウザ (Internet Explorer) の起動や終了，指定され

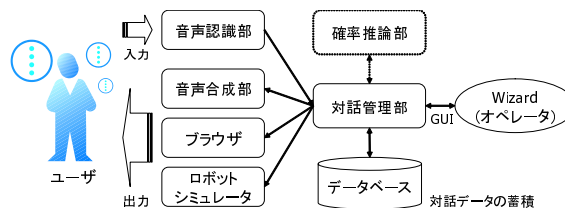


図 3: システム構成

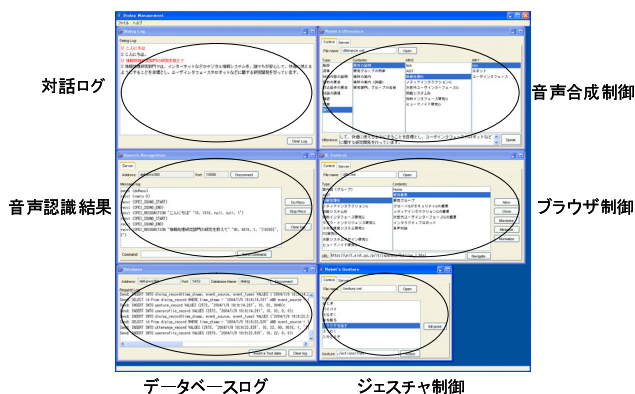


図 4: Wizard 用 GUI



図 5: ユーザ用画面の例

た Web ページの表示などを行うことが可能である。ロボットシミュレータは、[原 02] で開発されたソフトウェアの一部であり、PC の画面に 3D CG を用いてロボットを表示し、XML で記述されたスクリプトを用いて複合的なジェスチャを提示することができる。対話管理部は、人間のオペレータ (Wizard) が音声認識部や音声合成部、ブラウザ制御部、ロボットシミュレータの各モジュールを制御することで、ユーザとの対話を行い、対話データを収集するための GUI プログラムである。Wizard 用の GUI 画面を Fig.4 に示す。

3.2 動作の流れ

データ収集の流れは以下のようになる。

1. ユーザ用の PC 上で音声認識、音声合成、ブラウザ制御、ロボットシミュレータの各モジュールを、Wizard 用の PC 上で対話管理モジュールを起動する。また、対話管理部から音声認識部に認識依頼を行い、ユーザの発話があるまで待機する。
2. 音声認識部がユーザの発話を検出すると、認識結果を対話管理部に通知し、通知を受け取った対話管理部は、データベースにイベントを記録するとともに、Wizard 用の画面に認識結果を表示する。

3. Wizard は認識結果およびユーザの音声^{*1}からユーザの要求を推測し、画面上のリストから適切と思われるシステムの応答を選択し、ユーザに音声を出力する。また、このときにシステムの発話内容に関連する Web ページがある場合には同時に表示し、発話内容に対応するジェスチャ(たとえば「こんにちは」のときには“おじぎ”；「さようなら」のときは“手を振る”)がある場合には同時にジェスチャを実行させる。このときシステムが行った動作の内容も、自動的にデータベースに記録される。

4. ユーザが対話終了の音声(「ありがとう」や「さようなら」)を入力するまで、上記の 2 と 3 を繰り返す。

Fig.5 に対話中にユーザ用のディスプレイに表示される画面の例を示す。なお、対話中のユーザの様子はビデオで記録し、音声認識の正解率の計算などに利用した。また、対話終了後に簡単なアンケートを行い、対話の円滑さやシステムの有効性についての主観的評価を収集した。

4. 実験

4.1 対話データ収集

作成したシステムを用いて、被験者に対して対話実験を行い、対話データを収集した。システムが扱うタスクは、産業技術総合研究所の情報処理研究部門および知能システム研究部門の案内タスクとした。システムは、WWW サイト <http://www.aist.go.jp/> 上の情報も使いながら、ユーザに対して研究所の各研究グループの研究紹介や場所の案内などを行う。

ユーザプロフィール変数としては、駒谷ら [駒谷 03] と同様に、以下の二つを考えることにした。

- 対話内容に関する知識度 { 高い, 低い }
- 音声対話システムに関する習熟度 { 高い, 低い }

対話内容に関する知識度は、対話中のユーザが対話内容に関してどの程度の知識を持っているかを表す指標である。研究所の案内タスクの場合、ユーザは主に研究者が、それ以外の一般人に分けられ、対話内容に関する知識はユーザごとに大きな差がある。たとえば、研究内容の説明をする際に、一般人、つまり、対話内容に関して知識が低いユーザに対しては、専門用語を用いた詳しい説明をすることは避け、研究者のような対話内容に関して高い知識を持っているユーザに対しては、自明と思われるような情報は提供しない、というように、ユーザの知識に応じた適切な情報提供を行うことが望ましい。

音声対話システムに関する習熟度は、対話中のユーザの音声対話システムに対する慣れを表す指標である。ユーザが対話システムを使用する場合、システムが認識可能な単語や文法などがだいたい想像できると、対話が円滑に進む。しかし、現在のところ音声対話システムは一般的ではなく、ユーザの慣れには大きな差がある。たとえば、ユーザの習熟度が低い場合には、対話の主導権をシステムが取り、「お知りになりたい情報を次の中から選んでください」などとユーザの発話を促したりすることで、ユーザがタスクを放棄してしまうことを防止し、逆に、ユーザの習熟度が高い場合には、上記のような誘導はせず、対話の主導権をユーザに与え、自由な発話を許すことで、情報検索に必要な対話のターン数を減少させる、というように

*1 自由な発話に対する認識精度が十分ではないため、ユーザの音声 Wizard に聞こえるようにした。

表 1: 対話実験によって収集した各グループごとの対話パラメータの値

	ユーザの応答率	ユーザの平均応答時間 [s]	音声認識の成功率	システムの挨拶発話率	システムの確認誘導発話率	システムの案内発話率	システムの非有効発話率
グループ 1	0.86	2.95	0.52	0.24	0.17	0.34	0.25
グループ 2	0.85	4.44	0.43	0.26	0.08	0.51	0.13
グループ 3	1.02	3.70	0.40	0.26	0.03	0.43	0.27
グループ 4	0.91	5.89	0.42	0.25	0.07	0.40	0.20

適切な対話戦略をとることで、各ユーザに応じた効率の良い対話を実現することが望ましい。

プロフィールの推定に用いる対話パラメータについても[駒谷 03]を参考として、ユーザとの対話中に得られる情報(ここでは対話パラメータと呼ぶことにする)から抽出される、対話中のユーザの間合い、ユーザの一回の発話に含まれるキーワードや検索の絞込条件の数、音声認識の誤認識の割合(ユーザがシステムにあわせて発話をしてきているか)などを候補とした。

被験者は、産業技術総合研究所の職員や実習生などで、以下の4つのグループに分け、各グループ4名ずつ合計12名に対して行った。グループ1:対話システムとシステムが扱うコンテンツの両方に詳しい(習熟度:高,知識度:高),グループ2:対話システムを使ったことはあるがコンテンツはあまり詳しくない(習熟度:高,知識度:低),グループ3:対話システムを使ったことはないが、コンテンツはある程度知っている(習熟度:低,知識度:高)グループ4:どちらも知らない(習熟度:低,知識度:低)。被験者12名に対して、23対話を収集した。そのうち、10対話があらかじめ課題を設定した対話、13対話が課題を設定しない自由な対話である。

Table 1 に実験によって収集したデータから対話パラメータの候補となりえるものを抽出し、各グループごとに平均値を求めたものを示す。システムの案内発話率および非有効発話率は、対話中にシステムが行った全発話に対して、研究紹介などの発話を行った割合およびユーザの要求に対して有効な応答ができなかった割合を、それぞれ表している。この結果からは、平均応答時間や音声認識の成功率などが、対話パラメータとして有効である(論理的に考えて妥当でもある)と考えられる。

4.2 ユーザプロフィールの判別実験

対話実験で収集した対話データを用いて、ベイジアンネットワークによるユーザプロフィール(上記の4つのグループ)の判別実験を行った。ユーザプロフィールの判別を行うには、条件付確率 $P(C|P)$ を設定する必要があるが、ここでは、ナイーブベイズモデル(各対話パラメータが条件付独立と仮定)を用いて設定を行った。つまり、Table 1 の各対話パラメータを C_1, C_2, \dots, C_N とし、

$$P(C|P) \cong \prod_{i=1}^N P(C_i|P) \quad (1)$$

と近似することで、収集した対話データから条件付確率の推定を行った。また、本来は一発話ごとにプロフィールを判別すべきだが、ここでは、簡単のために一対話ごとのデータで判別を行い、23対話すべてのデータを学習データとしたときの closed な条件における判別精度を求めてみた。それぞれの条件付確率 $P(C_i|P)$ を正規分布と仮定して推定した場合には0.78、カーネル密度関数を用いて推定した場合には0.87の正答率であった。

5. おわりに

本研究では、ユーザにあわせて柔軟な情報提供を行うことが可能な対話システムを実現するためのアプローチを提案した。提案したアプローチを用いたシステムの実装に必要な基礎的データを収集するために、WOZ法による対話データ収集システムを作成した。実際にシステムを用いて予備的に対話データを収集し、収集したデータからベイジアンネットワークを用いてユーザプロフィールの判別実験を行い、その有効性を検証した。

今後、実際の対話システムに組み込むためには、より多くの対話データを収集して、さらに判別精度を改善することが必要である。今回の実験では、対話タスクの指定のしかたが適切ではない面があり、それによって対話の様子がかなり異なってしまうということが観察された。従って、本格的な対話実験を行う際には、タスクの設定や、被験者に対する対話タスクの指定のしかたについてさらに検討することが必要である。

謝辞: 本研究の一部は科研費 14208033 による。

参考文献

- [加藤 95] 加藤 恒昭: 対話システム, 田中穂積(監修)「自然言語処理 - 基礎と応用 -」第9章, pp.281-381, (社)電子情報通信学会(1995).
- [鹿野 02] 鹿野 清宏: 音声対話機能を持つ受付案内ロボット ASKA の実装と評価, 言語理解と行動制御 平成13年度研究成果報告書, pp. 149-161 (2002).
- [松阪 02] Matsusaka, Y., Tojo, T., and Kobayashi, T.: Conversation robot participating in group conversation, *IEICE Trans. on Information and Systems*, vol.E86-D, pp.26-36 (2003).
- [駒谷 03] 駒谷 和範, 上野 晋一, 河原 達也, 奥乃 博: ユーザモデルを導入したバス運行情報案内システムの実験的評価, 情報処理学会研究会資料, SLP-47-12, pp. 59-64 (2003).
- [Russel 95] S. Russell and P. Norvig: Probabilistic Reasoning System, In *Artificial Intelligence, A Modern Approach*, pp.436-470, Prentice Hall (1995) (2nd Edition 2002), (古川 康一 監訳: 「エージェントアプローチ人工知能」, pp. 439-473, 共立出版(1997)).
- [本村 03] 本村陽一: ベイジアンネットワークによるヒューマンモデリング, 2003年ベイジアンネットワークセミナー(BN2003) 予稿集, pp.79-84 (2003).
- [麻生 03] 麻生 英樹, 小玉 智志, アブデラジズ・キアット, 松本 泰明, 本村 陽一, 原 功, 浅野 太, 新田 恒雄, 小笠原 司, 柿倉 正義: “確率的推論を利用したマルチモーダル対話制御”, 2003年度人工知能学会全国大会(第17回) 論文集, 1C1-04 (2003).
- [Julius 04] “大語彙連続音声認識システム Julius”, <http://julius.sourceforge.jp/> (2004).
- [原 02] 原 功, 本村 陽一, 麻生 英樹, 河村 進: インタラクティブ・ロボット基本ソフトウェアの開発, 平成13年度未踏ソフトウェア創造事業開発成果論文(2002).