1O2-OS-18-11

POMDP を用いた聞き役対話制御部の Wizard of Oz 実験による評価

Wizard of Oz experiment of listening-oriented dialogue control using POMDPs

目黒 豊美[*1] 南 泰浩[*1] 東中 竜一郎[*2] 堂坂 浩二[*1†]
Toyomi Meguro Yasuhiro Minami Ryuichiro Higashinaka Kohji Dohsaka

[*1]NTT コミュニケーション科学基礎研究所, 日本電信電話株式会社
NTT Communication Science Laboratories, NTT Corporation

[*2]NTT サイバースペース研究所, 日本電信電話株式会社
NTT Cyber Space Laboratories, NTT Corporation

We have been working on dialogue control for listening agents and have proposed a dialogue control method that maximizes user satisfaction using partially observable Markov decision processes (POMDPs) [Meguro 10]. Although our method significantly outperformed other stochastic dialogue control methods in simulation, this does not necessarily mean that our method works as well in real dialogues with human users. In this paper, we evaluate our dialogue control method by a Wizard of Oz (WoZ) experiment. The experimental results show that our POMDP-based method achieves significantly higher user satisfaction than other stochastic models, confirming the validity of our approach.

1. はじめに

本稿では，対話において一方が聞き役となり，もう一方の話し役の話を聞くような対話を「聞き役対話」，人の聞き役となる対話システムを「聞き役システム」と呼ぶ．我々は，ユーザの話に積極的に耳を傾け「話したいという欲求」や「聞いてもらいたいという欲求」を満足させることで，ユーザの心理状態を良好な状態に導く「聞き役システム」の構築を目指している．図 1 は典型的な聞き役対話の例である．本研究では聞き役システムの対話制御部に着目する．対話制御部は，発話制御部が出力したユーザの対話行為をもとに，次の最適なシステムの対話行為を決定する部で，対話の流れを決める最も重要な部の一つである [Meguro 09].

従来，聞き役システムのように，達成すべきタスクが明確でない対話システム（非タスク指向型対話システム）における対話制御には人手でルールを記述していた．しかし，タスクによる制約が少ない状況下において，ルールベースの対話制御では，すべての対話状態に対応したルールを書ききることは難しい．この問題を解決するためには，データから対話制御部を自動構築する手法が考えられる．我々は非タスク指向型対話への POMDP の応用を提案してきた [Meguro 10]．しかし，聞き役システムが行うような非タスク指向型対話では，タスク指向型のようなはっきりとしたユーザ目的を定義できず，妥当な報酬関数を設定することが難しい．

我々の手法では，POMDP を聞き役対話に適用するために，「ユーザがシステムに話を聞いてもらえていると感じているか (以後，ユーザ満足度)」と「システムが自然な対話を生成できているか (以後，自然性)」を最大化するポリシーをデータから学習する．そのためにまず，学習のために大量の聞き役対話を収集し，対話行為タグと対話満足度（主観評価）を付与する．そして，そのデータから報酬を計算し，POMDP のポリシーを学習する．その後，学習したポリシーを用い対話行為タグを生成する対話制御部を構築する．[Meguro 10] ではシミュレーションによる評価を行ったが，実際にユーザがシステムを使うと異なる評価になる可能性があり，その有効性が検証されたとは言えない．そこで，本稿では，Wizard of Oz (WoZ) 形式で実験参加者による主観評価を行う．

発話	対話行為
S: こんばんは	挨拶
L: こんばんは	挨拶
S: テーマは食事でお願いします	挨拶
今夜の夕飯はカレーでした	自己開示-事実
L さんはカレーは好きですか？	質問-評価
L: よろしくお願いします。	挨拶
何カレー？	質問-事実
S: 自宅カレーです	自己開示-事実
L: カレー大好きです!	共感・同意
S: 隠し味などナッスィングなカレーです	自己開示-事実
L: ナッスィングなカレーですね！	繰り返し

図 1 典型的な聞き役対話．対話行為は一文章ごとにアノテートされている．S は話し役，L は聞き役.

2. 関連研究

聞き役対話を扱った先行研究として Maatman らの研究 [Maatman 05] が挙げられる．この研究では，バーチャルエージェントのジェスチャーや，頷き，頭部の動きによってユーザに「聞いてもらえている」という感覚を与えている．これに対して，我々は言語的に「聞いてもらえている」という感覚を与えることを目的としている．下岡らの研究 [下岡 10] では，聞き役の返答生成に着目し，音声認識信頼度を用いて，高い信頼度の時には「繰り返し/問い返し発話」または「共感」を，低い信頼度の時には，「相槌」を行う．また，横山ら [横山 10] は，対話エージェントとのインタラクションを長く保持させるために，傾聴モードと話題提示モードを切り替える手法を提案している．これらのシステムはルールを用いて対話を制御しているが，我々の目的は，ユーザに「聞いてもらえている」と感じてもらえるシステムを自動的に対話データから学習し構築することである．

Williams らは，POMDP をタスク指向型対話 (チケット購

連絡先: meguro.toyomi@lab.ntt.co.jp
† 現在，秋田県立大学

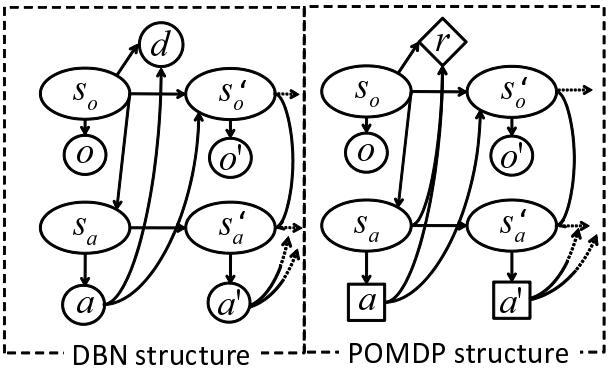

図 2　DBN と POMDP の構造

入タスク) に適用している [Williams 07]．これに対して，本稿は，聞き役対話のようにユーザの目的がはっきりとしていない対話への適用を行う．我々はユーザの目的のはっきりしない対話，すなわち，非タスク指向型対話に POMDP を適用することに取り組んできており，これまで，シミュレーションで生成したエージェントとユーザの疑似データから対話制御部の学習を行なってきた [Minami 09]．これに対して，本稿では，実際の人同士が行った聞き役対話データから対話制御部を学習する．

3. POMDP を用いた聞き役対話制御

本節では，我々の提案方法を説明したあと [Meguro 10]，本稿で加えた改良点について述べる．

3.1 聞き役対話のためのポリシーの学習方法

POMDP を構築する前に，統計的構造を得るため収集した大量の聞き役対話から DBN を学習する．DBN の確率変数は次のように設定した．s_o は対話状態，s_a はアクションの状態，o は話し役の観測値，a は聞き役のアクション，d はユーザ満足度の評価値の変数である．評価値は，各アクションについてアンケートから得られる値で，POMDP でユーザ満足度を計算するために使われる．図 2 内の DBN の矢印は出力確率と遷移確率を表現している．矢印から，o' の出力確率は $Pr(o\,'|s_o\,')$ であり，d の出力確率は $Pr(d|s_o)$ であり，s_o から s'_o への遷移確率は $Pr(s_o\,'|s_o, a)$ と記述される．DBN のこれらの確率値は EM アルゴリズムにより学習される．

図 2 の左の DBN から右の POMDP へと変換する．r は満足度に対する報酬 $r1$ と自然性に対する報酬 $r2$ の和で，$r1$ は平均満足度，r_2 はアクションの予測確率が大きくなるように設定された報酬である．時刻 t の時点で将来得られる平均報酬を下の式によって計算する．

$$V_t^{\pi} = E^{\pi}\left[\sum_{\tau=0}^{\infty} \gamma^{\tau} r((s_o, s_a), a_{\tau+t})\right] \tag{1}$$

τ は減衰関数で，将来的な報酬は τ によって減衰する．ポリシーは，V_t の平均を最大にするアクションが選ばれるように value iteration [Pineau 03] によって学習される．

3.2 改良 1: 複数対話行為の許容

図 2 の POMDP の構造では，システムの対話行為と，ユーザの対話行為は必ず 1 ターンにひとつずつでなければならない．しかし，図 1 の対話例のように，人同士の自然な対話において，1 ターンに複数の対話行為が含まれることが多い．聞き役対話の本来の目的は，ユーザが自由に話し，「聞いてもらえた」という感覚を起こすことであり，1 ターンにつき，1 つの対話行為に限定することは，本目的に反することになる．また，システムの発話が 1 つの対話行為に限定されてしまうのも不自然であろう．

そこで，「スキップ」という対話行為を導入することにより，1 ターンで複数の対話行為を扱えるように改良する．収集した聞き役対話中で，1 ターン中に複数の対話行為がある場合には，対話行為間に対話相手が「何も発話しないという対話行為」を実施したと考え，「スキップ」を挿入する．その上で，3.1 節の通りにポリシーを学習する．

具体的には，ユーザが 1 ターン中に複数の対話行為を入力した場合には，対話行為間にシステムが「スキップ」をしたと見做し，出力確率と遷移確率を計算する．一方，システムのターンの時には，次のユーザの対話行為を図 2 の POMDP を使って予測する．ユーザの次の対話行為が「スキップ」であると予測されれば，システムはその次のシステムの対話行為の計算を続ける．この時点で，対話行為の出力はまだ行わない．逆に「スキップ」以外であると予測された場合には，システムは計算を止め，それまでに選択されたシステムの対話行為を出力する．システムは常に報酬が最大になる対話行為を選択するが，例外的に，システムのターンの最初の対話行為が「スキップ」となってしまった場合，そのままではユーザの発話を無視することになるので，2 番目に報酬を最大にする対話行為を選択する．

3.3 改良 2: システムアクションによるフィードバック

実際のシステムが動作する時は，ポリシーに従い最も適切な対話行為を選択する．このとき，その対話行為はシステムにとって既知となる．この情報をシステム側にフィードバックすることによって，より適切な次のシステムの対話行為を選択することがと期待できる．

図 2 に示す POMDP では s_a の状態数は対話行為数と同じでひとつずつ各対話行為に対応している．次のアクションの予測確率を求めるため，ここでは，隠れ状態を確率分布として扱う．この確率分布にしたがい次のアクションは選択されるが，アクションが選択された後では，状態の分布を確定的に扱うのは当然だと考えられる．しかし，以前の対話制御 [Meguro 10] では，これを確率的に扱っていた．これを確定的に扱うため，ここでは，システムが，自身の対話行為を選択した直後に，s_a の確率分布にアクションの出力確率を反映させることで，アクションが選択された情報を状態にフィードバックする．

4. 実験

4.1 実験設定

我々の提案手法 [Meguro 10] に 3.2 節，3.3 節の改良を加えたシステムを，WoZ 実験によって評価した．実験では，Wizard が聞き役，評価者が話し役となり，別々の部屋でテキストのみを用いて会話した．声，画像は伝わらず，顔文字などは禁止とした．比較手法として，4.2 節で説明する 6 システムを用意し，提案手法と合わせて計 7 つの対話制御システムを，以下の方法で比較した．

1. 話し役として評価者は，チャットシステムに自然言語を用いてシステムに話しかける．話し役は，システムの裏に人 (Wizard) がいることは知らない．
2. Wizard は評価者の発話を見て，適切な対話行為に変換する．そして，評価者の発話に含まれる対話行為列を対話制御システムに入力する．
3. システムは入力された対話行為列を基に次のシステムの対

話行為列を決定し，出力する．

4. Wizard はシステムが出力した対話行為列を基にテンプレートを用いて，システムの発話を生成する．テンプレートは筆者が作成したもので，各対話システムにひとつ以上のテンプレートが用意されている．一部のテンプレートは空欄があり，Wizard は話の流れに沿って，指定されたカテゴリの単語の中から適切なものを選択し，空欄を埋めることができる．
5. Wizard はシステム発話をチャットシステムに入力し，次の評価者の発話を待つ．(ステップ 1 に戻る)

実験参加者は評価者 14 人と Wizard 2 人 (それぞれ男女同数, 著者は含まない) で, 休憩中も含め, 顔を合わせることがないように別々の部屋で実験に参加した．評価者は 2 人の Wizard とそれぞれ 7 回ずつ (1 システムにつき 1 回) の計 14 回対話を行った．対話時間は 15 分で，テーマは「食べ物」とした．各対話後には「対話システムはいい聞き役だったと思いますか？」という項目を 7 段階評価を行った．同様に，Wizard は「対話全体を通して，システムの出力したタグ (対話行為) は聞き役としてよかったと思いますか？ 」という項目について，7 段階で評価した．

4.2 システム

提案方法に加えて，6 つの比較用システムを用意した．POMDP と HMM の学習には，先行研究 [Meguro 10] で収集済みの人同士の聞き役対話コーパスを用いた．聞き役対話コーパスには 1259 対話の各文に対話行為が人手で付与されている．

提案システム　このシステムは我々の提案手法を用いたシステムで，状態数，2 つの報酬の重み係数は [Meguro 10] と同じである．

満足度のみ POMDP　我々が提案した二つの報酬がそれぞれどのような効果があるか検証するため，一つの報酬ずつしか使わないシステムを用意した．このシステムは，満足度の報酬のみを用いたシステムで，報酬以外の設定は提案システムとまったく同じである．

自然性のみ POMDP　このシステムは，満足度のみ POMDP と同様に，自然性の報酬のみ用いたシステムである．

ルールによる制御　このシステムは筆者が書いたルールに基づき制御を行う．ルールは，我々の聞き役対話の分析 [Meguro 09] と先行研究をベースに作成されたもので，先行研究において，その有効性が示されている [Meguro 11]．

HMM　このシステムは聞き役対話から学習した HMM を用いて最尤な対話行為を選択する制御を行う．確率的でなく最尤な対話行為を選択する方式にした理由は，HMM を用いて確率的に対話制御を行う手法も考えられるが，ここでは，可制御性と制限性を考慮に入れて，最尤な対話行為を選択する手法を取った．本システムは，提案手法と同様に複数対話行為を扱うために，次の対話行為の推定を行う．具体的には，次の対話行為がシステムのものであった場合には，システムの対話行為をスタックし，次の対話行為がユーザのものであった場合には，推定を止め，システムの対話行為列として，スタックしていたものを出力する．HMM の状態数は 18 とした．

人による制御　このシステムは Wizard が自由に対話行為を選択する．

ランダムシステム　このシステムでは，すべての対話行為を等確率でランダムに出力する．対話行為数は 1〜4 の間から，ランダムに選択する．

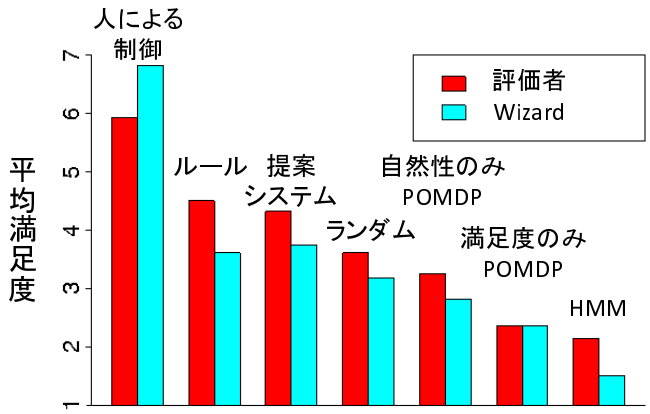

図 3　評価者と Wizard がつけた平均満足度．1〜7 の 7 段階で評価．

4.3 評価結果

提案方法と比較手法を合わせた計 7 つのシステムに対する主観評価の平均値は図 3 のようになった．赤で表した「評価者の満足度」とは話し役を行った評価者による対話に対する評価値，青で表した「Wizard の満足度」とは聞き役を行った Wizard がシステムが出力した対話行為列を評価した値である．

我々の提案手法は，ルールによる制御とほぼ同等の性能であり，ルール以外のシステムよりは有意に満足度が高いということがわかった ($p<0.05$, t-test)．提案手法は，一つの報酬しか使っていない POMDP より満足度が高かった．つまり，報酬は二つとも必要であるということを示している．また，統計的な制御方法として最も代表的な HMM より満足度が高かったのは，確率だけに基づく制御は不十分であるということを示している．ルールによる制御とはほぼ同等の満足度だが，ルールは，専門家が，収集した対話を数か月間分析を行って作成したものである．一方，我々の手法では，単純にデータから学習しているだけで同様の効果が得られている．

ランダムシステムは，HMM や報酬を一つしか使わない POMDP より満足度が高かった．理由として，一発話における対話行為数が考えられる．HMM と POMDP の学習に使った聞き役対話を調べてみたところ，聞き役の対話行為は 1 ターン中，平均 1.6 個の対話行為が含まれていた．そのため，HMM と POMDP も同様にほとんどの場合 1 ターン中に 1 個または 2 個の対話行為しか出力しなかった．しかし，今回の実験で使ったチャットインタフェースは聞き役対話を収集した時のものとは異なり，より長い文章がが入力されやすい仕様となっていた．実際に，評価者が入力した一ターンあたりの平均対話行為数は 2.4 個であった．ランダムシステムは対話行為数が 1〜4 の中からランダムに選択するようにしており，評価者の平均対話行為数と近い値になっていた．そのため，他のシステムよりランダムシステムの方が出力する発話の長さが，評価者の入力する長さに近かった．この人間の入力とシステムの入力の長さが同じという事実は，ランダムシステムに対し，人間との対話に近い印象をを与えた可能性がある．ランダムシステムの満足度が高いもう一つの理由として，テンプレートの自由度が考えられる．上で述べたように，ランダムシステムは 1 ターンで 1〜4 個の対話行為を出力する．ここで，使用した対話行為セット中，空欄を含む対話行為 (質問，自己開示) は約半分を占めていることを考えると，1 ターン中に空欄のあるテンプレートを持つ対話行為がほぼ必ず一つは含まれることになる．このため，Wizard

発話・対話行為 (5-best)
L: はじめまして [挨拶:26.0 質問-欲求:24.9 承認:20.7 スキップ:20.2 自己開示-評価 N:20.1]
S: よろしくお願いします [挨拶]
L: 海ぶどうを食べてみたいですか？ [質問-欲求:24.0 スキップ 21.9 挨拶:21.1 承認:20.1 自己開示-評価 N:20.1]
S: 食べてみたいですね。 [共感] 食べたことはありませんね。[自己開示-経験]
L: 沖縄で食べるのが好きです [スキップ:25.6 自己開示-評価 P:21.0 確認:20.7 自己開示-欲求:20.7 情報提供:20.4]
S: 今回のテーマは旅行先での食べ物でお話させてください。 [自己開示 事実] わたしも沖縄には思い出があります。 [自己開示-評価 P]
L: そうなんですか？ [スキップ:24.1 確認:20.5 自己開示-評価 P:20.5 情報提供:20.0 質問 事実:20.0]
S: ラフテーやゴーヤチャンプルなんかを食べましたね。 [自己開示-事実]

図 4　提案手法を用いた対話例．括弧内は，提案システムが報酬が高いと出力した対話行為の 5-best と，その報酬値．“評価 P” はポジティブな評価，“評価 N” はネガティブな評価を指す．S は話し役 (評価者)，L は聞き役 (Wizard)．

発話	対話行為
L: はじめまして。	挨拶
こんにちは	挨拶
S: こんにちは。	挨拶
よろしくお願いします。	挨拶
外食についてお話させてください	挨拶
L: はい。	共感
よろしくお願いします	挨拶
S: わたしはあまり外食はしません。	自己開示-事実
L: 本当ですか？	確認
自宅で食べることが多いですか？	質問-習慣
S: そうですね。	共感
自宅が多いです。	共感
外食はよくしますか？	質問-習慣
L: はい。	共感
私は定食屋で食べることが多いです。	自己開示-習慣

図 5　Wizard が自由に評価者と対話を行った時の対話例．

の裁量が大きく，話しの流れに合わせやすかった可能性がある．例えば，Wizard は，例え対話の流れとして不自然な対話行為が出力されたとしても，修正がしやすかったのではないかと思われる．

Wizard による評価では，ルールシステムとランダムシステムとの間を除いて提案手法の満足度が有意に高かった ($p<0.05$, t-test)．評価者による満足度との順序を比較すると，ルールシステムと提案方法が入れ替わっている以外同じであった．Wizard は直接対話行為を見て評価を行っているため，テンプレートの特徴による影響は受けていない．つまり，提案手法はルールと比較し，対話行為による制御では，よりよい制御を行えているということを示している．

4.4　実験で得られた対話

図 4 と図 5 は，提案手法を用いた時と，Wizard が自由に対話を行った時の対話である．図 4 では対話結果と共に，提案手法の POMDP が出力した報酬の高い対話行為を各 5 つ表示している．この結果から，提案手法は対話の初めに挨拶を行い，対話が進むごとに，質問や自己開示を適宜行っていることがわかる．また，Wizard が自由に対話を行った図 5 と比較すると，“確認” を行って話し役が聞き役の話を理解し興味を持っていることを伝えたり，相手に質問することでユーザの話を掘り下げようとしたりする点が類似しており，提案方法が，適切な制御が行っているということがわかる．

5.　おわりに

本稿では，POMDP を用いた対話制御方法を Wizard of Oz (WoZ) 実験によって評価し，提案手法が他の統計的手法より満足度が有意に高いことを示した．タスク達成の代わりにユーザの満足度を最大にするような報酬を学習した時においても，POMDP を用いた対話制御が有効であることがわかった．

今後，報酬として満足度と自然性を用いたが，他の適切な報酬がないか検討したい．我々は，対話制御部の学習に用いるために人同士の聞き役対話を大量に収集したが，これは大変コストが高い．そこで，Twitter データなど自動的に集められる対話を用いてより自然な生成が行える対話制御を検討したい．加えて，今回扱った対話制御は，対話行為しか扱えない．今後は，自然言理解部と生成部を自動化することによって，人の手を介さず応答を生成できる，満足度の高い聞き役システム開発を目指していく予定である．

参考文献

[Maatman 05] Maatman, R. M., Gratch, J., and Marsella, S.: Natural behavior of a listening agent, *Lecture Notes in Computer Science*, pp. 25–36 (2005)

[Meguro 09] Meguro, T., Higashinaka, R., Dohsaka, K., Minami, Y., and Isozaki, H.: Analysis of Listening-oriented Dialogue for Building Listening Agents, in *Proc. SIGDIAL*, pp. 124–127 (2009)

[Meguro 10] Meguro, T., Higashinaka, R., Minami, Y., and Dohsaka, K.: Controlling listening-oriented dialogue using partially observable Markov decision processes, in *Proc. COLING*, pp. 761–769 (2010)

[Meguro 11] Meguro, T., Higashinaka, R., Minami, Y., and Dohsaka, K.: Evaluation of Listening-oriented Dialogue Control Rules based on the Analysis of HMMs, in *Proc. Interspeech*, pp. 809–812 (2011)

[Minami 09] Minami, Y., Mori, A., Meguro, T., Higashinaka, R., Dohsaka, K., and Maeda, E.: Dialogue Control Algorithm for Ambient Intelligence based on Partially Observable Markov Decision Processes, in *Proc. IWSDS*, pp. 254–263 (2009)

[Pineau 03] Pineau, J., Gordon, G., and Thrun, S.: Point-based value iteration: An anytime algorithm for POMDPs, in *Proc. International Joint Conference on Artificial Intelligence (IJSAI)*, pp. 1025–1032 (2003)

[Williams 07] Williams, J. and Young, S.: Partially observable Markov decision processes for spoken dialog systems, *Computer Speech & Language*, No. 2, pp. 393–422 (2007)

[横山 10] 横山 祥恵, 山本 大介, 小林 優佳, 土井 美和子：高齢者向け対話インタフェース—雑談継続を目的とした話題提示・傾聴の切替式対話法—, 情報処理学会研究報告 音声言語情報処理, SLP-80, No. 4, pp. 1–6 (2010)

[下岡 10] 下岡 和也, 徳久 良子, 吉村貴克, 星野博之, 渡部生聖：音声対話ロボットのための傾聴システムの開発, 人工知能学会研究会資料, SIG-SLUD 58, pp. 61–66 (2010)