

POMDP を用いた聞き役対話システムの対話制御

目黒豊美[†], 東中竜一郎[‡], 南泰浩[†], 堂坂浩二[†]

[†] 日本電信電話株式会社, 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 サイバースペース研究所

{meguro.toyomi, higashinaka.ryuichiro, minami.yasuhiro,
dohsaka.kohji}@lab.ntt.co.jp

1 はじめに

従来, タスク指向型の対話システムが盛んに研究されてきた [4]. しかし近年ではタスクを重視せず, チャットのような雑談に近い対話システムの社会性やエンターテインメント性の機能が注目されてきている. また, カウンセラーや傾聴ボランティアの重要性が認識されてきている. このような状況を踏まえ, 本研究では対話参加者の一人がもう一方の参加者の話を積極的に聞く対話 (以後, 聞き役対話) を扱う. 本研究の目的は, ユーザの話を聞くことによって「話したい」「聞いてもらいたい」というユーザの欲求を満たすシステムの構築である. 図 1 は典型的な聞き役対話の例である.

本システムは対話全体を通して「聞いてもらえた」という感覚を与えることを目的としており, そのような感覚を与えるためには, 対話制御, 言語生成の両面を考慮する必要があるが, 今回は, 特に対話制御に着目する. なぜなら, 対話制御は話の流れを決定するより大きな要素と考えられるからである.

従来, 非タスク指向型の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, ルールベースの対話制御では, タスクによる制約が少ない状況下において, 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, 本研究では部分観測マルコフ決定過程 (POMDP) の枠組みを用いて, 平均獲得報酬を最大化させるように, 聞き役エージェントの対話制御部をデータから自動構築する. 聞き役エージェント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く, 非タスク指向型対話への POMDP の応用例はない.

本手法では, 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

発話	対話行為
S: こんにちは 「食事」をお願いします	挨拶 挨拶
L: はい, よろしくお願いします	挨拶
S: 今日の夕飯はカレーでした Bさんはカレーは好きですか?	自己開示 (sub: 事実) 質問 (sub: 評価)
L: 好きです.	共感・同意
S: おお, お好きですか. 私も好きなんです.	繰り返し 共感・同意
L: 外食が主ですか?	質問 (sub: 習慣)
S: いえ, 自宅で作ります. 特に隠し味はありませんが, かつおだしでのばしてうどんに もします	自己開示 (sub: 習慣) 自己開示 (sub: 習慣)
L: うわー, それすごい美味し そうです!	自己開示 (sub: 評価 (ポジティブ))

図 1: 典型的な聞き役対話例. 対話テーマは「食事」で, 一文毎に対話行為タグが一つずつ付与されている. S は話し役, L は聞き役.

最大化するポリシーをデータから学習する. はじめに学習のための大量の聞き役対話を収集し, 対話行為タグと主観評価による対話満足度を付与した. 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列をシミュレーションで生成し, 実験参加者による主観評価を行った.

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 [1] があげられる. 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顔つき, 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 [6] では, 聞き役の返答生成に着目し研究している. この研究では音声認識結果から信頼度を判定し, 高い信頼度のときには「繰り返し/問い返し発話」または「共感」を行い, 低い信頼度のときには,

「相槌」を行う。また横山らの研究 [7] 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。これらのシステムはルールを用いて制御しているが、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。

POMDP は Williams らが、タスク指向型対話 (チケット購入タスク) に適用している [5]。これに対して、本稿で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。先行研究として他の非タスク指向型対話システムへの適用がある [3]。この研究では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シミュレータを学習している。本稿では、実際の人同士の行った聞き役対話データから学習しているという点で異なる。

3 POMDP を用いた対話制御

先行研究 [2] で聞き役対話で聞き役は積極的に質問を行うだけでなく、間に自己開示を挟むことによって、社会的関係を構築しようとしていることがわかっている。このような聞き役の特徴的な流れを実現する対話システムの対話制御部を自動的に構築するため、統計的に POMDP の報酬とポリシーを大量の聞き役対話データから学習する。POMDP によって将来的に得られる報酬を最大化するアクション系列を選択するポリシーを学習することができる。そのため、報酬の設定が POMDP の中では最も重要である。

本稿では満足度と自然性の 2 つの報酬を提案する。POMDP を構築する前に、統計的構造を得るため 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、アンケートから得られた値で、変数は 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。 $Pr(o' | s_o)$ は o' の出力確率、 $Pr(d | s_o)$ は d の出力確率、 $Pr(s_o' | s_o, a)$ は s_o から s_o' への遷移確率である。DBN は EM アルゴリズムを用いて学習した。得られた変数を用い、二つの報酬を以下のように計算した。

(1) ユーザ満足度 : この報酬は、変数 d から次の式を用いて得られる。

$$r_1((s_o, *), a) = \sum_{d=\min}^{\max} d \times Pr(d | s_o, a),$$

*は任意の s_a 、 \min, \max はそれぞれ評価値の最小値と最大値である。

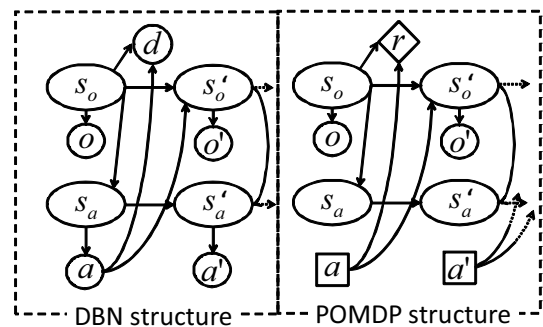


図 2: 提案手法の DBN と POMDP の構造。POMDP 内の a が独立しているのは、学習したポリシーによって決定されるためである。

(2) 自然性 : 本稿では自然性を高めるために、アクションと観測値の履歴から得られるアクションの予測確率を最大化する報酬 r_2 を用いる。 S_a を用いてアクションの予測確率を見積り、一定の r_2 をかけることによって、予測確率を最大にするアクションを選択することができる。

これら二つの報酬を用いて満足度と自然性の高いアクションを選択するポリシーを得るための目的関数を構成する。

3.1 POMDP から DBN への変換

POMDP を DBN に変換する (図 2)。遷移確率と出力確率を用いて、上記 (1) のように d から r に変換する。システムは部分観測状態にある。つまり、状態は一意に求まるものではなく、belief state b_t と呼ばれる確率分布で表わされる。確率分布を用い、時間 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。

$$V_t = \sum_{\tau=0}^{\infty} \gamma^{\tau} \sum_s b_{\tau+t}((s_o, s_a)) r((s_o, s_a), a_{\tau+t}),$$

τ は減衰関数で、将来的な報酬は τ によって減らされる。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 によって学習される。 $r((s_o, s_a), a)$ は以下のように定義した。

$$r((s_o, s_a), a) = r_1((s_o, *), a) + r_2((* , s_a), a)$$

この二つの報酬のバランスをとることで、満足度が高く自然性の高いアクションを選択できるようになる。

4 評価実験

4.1 対話データ

聞き役 10 人 (男女比同) と話し役 37 人 (男性 18 人女性 19 人) の実験参加者による聞き役対話を収集した。実験参加者は 20 代から 60 代の日本語母語話者で、聞き役と話し役にわかれ、インターネット上のチャット

表 1: 収集データの統計情報

対話数	1260
一対話あたりの平均発話数	28.2
聞き役一人当たりの対話数	126
話し役一人当たりの対話数	34
総対話行為数	67801
アノテーションの一致率	0.57

システム上で対話を行った。対話ではテキストのみを用い、音声、動画像、顔文字等の使用は禁止した。聞き役には、「話し役が話しやすいようにふるまうこと」とインストラクションを与え、計 1260 対話を収集した。データの情報は表 1 のようになった。

次に収集した対話データに 32 個の対話行為タグを二人のアノテータが一文につき対話行為タグを一つずつラベル付けを行った。一発話中に複数の文章が発話されることがあるため、ひとりのアノテータがまず発話を一文ごとにわけ、その後二人のアノテータが独立してそれぞれの文にラベル付けした。また、対話参与者ではない第三者がユーザ満足度を付与した。それぞれの対話において、話し役が「聞いてもらえた」と感じるか 7 段階の Likert 尺度で評価した。対話全体にひとつの満足度を評価したため、その評価スコアがその対話内のそれぞれのアクションに与えられたと設定し、POMDP を学習した。

4.2 実験手法

実験は 3 つの手順で行う。

(STEP1) 前述の提案手法の POMDP と 5 つの比較用システム (後述) を構築し、それぞれから評価用の対話行為列をシミュレーションによって出力する。学習用データは収集データの中で最も多かった「食事」の対話のみを用いた。これは、まずひとつのテーマで検証するためである。

(STEP2) 評価用の対話行為列を見て、直接主観評価を行うことは難しいため、一度自然な発話文に直し、評価する。自然文に直す際には、指定された日常的と思われる状況を設定し、対話を作成する。この対話作成は、19～39 歳の 16 人の実験参加者 (男女比同) が 6 システムの出力それぞれに 2 対話ずつ (計 12 対話) 作成した。

(STEP3) STEP2 とは異なる実験参加者を 3 人 (男性一人、女性二人) がそれぞれ STEP2 で作成されたすべての対話について「もしあなたが話し役だった場合、いい聞き役であると感じたか」という軸で 7 段階で評価を行う。

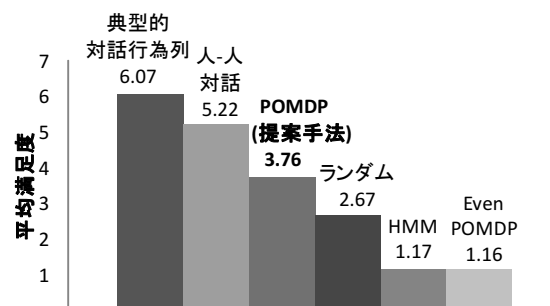


図 3: システムの平均満足度

4.3 評価用システム

提案手法を用いたシステムと、5 つの評価用システム計 6 つのシステムを以下のように構築した。

(提案手法)POMDP 提案手法によりポリシーを学習した。ユーザ観測値はシミュレータの出力確率に、システムアクションはポリシーに従い生成される。アノテータに使用した 32 対話行為タグに、スキップを加えた 33 タグを使用した。スキップは、聞き役と話し役の対話行為が交互になるようにするために使用する。 S_o と S_a の状態数はそれぞれ 16, 33, $r_2((*, s_a), a)$ は 10 とした。

(1)EvenPOMDP このシステムは満足度は使わず、自然性の報酬のみ使用した POMDP である。その他の条件は提案手法の POMDP と同じである。このシステムの目的は、満足度が報酬として必要がどうか評価するためである。

(2)HMM 先行研究 [2] の SpeakerHMM を用いる。システムアクションは最も高い確率のもの、ユーザアクションはユーザの対話行為の確率分布に従いランダムに選択する。

(3) 典型的な聞き役対話の系列 この系列は先行研究で得られた知見をもとにルールを作成し、ユーザ、システム両方のアクションがルールによって生成される。

(4) 人-人対話行為列 この系列は収集した人同士の対話データの中に実際にあった対話行為列を抽出してきた系列である。

(5) ランダム ランダムに対話行為列を生成する。

4.4 評価結果

主観評価は図 3 のようになった。HMM と POMDP 間を除き、すべてに有意な差 ($p < 0.01$) があった。この結果から、他の統計的手法を用いたシステムよりも提案手法がより「聞いてもらえた」と感じられるシステムが構築できていることがわかる。

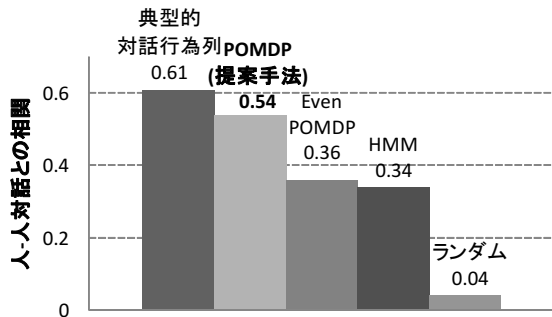


図 4: 人-人対話の対話行為の分布との相関

発話文	対話行為
L: 今日の夕飯はどこで誰と食べましたか?	質問 (sub: 事実)
S: 今日は自宅で家族と肉じゃがを食べました	自己開示 (sub: 事実)
L: ありがとうございます 何人兄弟ですか?	感謝 質問 (sub: 事実)
今度、兄夫婦が遊びに来るんです	自己開示 (sub: 予定)
今度、兄夫婦と一緒に肉じゃがを作るんです	自己開示 (sub: 予定)
S: そうなんです	共感・同意

図 5: POMDP が出力した対話行為列から作成した自然文作成例

対話例を図 5 に示す。聞き役は質問や自己開示を行い、先行研究で見られた聞き役対話の分析結果に見られた聞き役の典型的な行動がある程度再現され、話し役の話を引き出していることがわかる。しかし、3 発話目のように不自然に「感謝」の対話行為を生成してしまうということも見られた。これは、対話に含まれるすべての対話行為に対して一律に同じ評価値を与えているため、本来は「いい対話行為」ではなくても、高い評価を得られている場合があり、起こってしまった問題であると考えられる。

人-人対話における対話行為タグの出力分布の類似性をタグの出現確率の相関係数で求めたところ、図 4 のようになった。最も高いのは典型的な対話行為列で、人-人対話を適切に再現したルールを用いたことが、主観評価の高さにもつながったと思われる。次に相関のある POMDP は他の統計的手法に比べ、人間の対話行為の出力分布を再現できていることがわかる。

5 まとめ

我々は、ユーザの話を聞くことによって「話したい」という欲求を満たす聞き役対話システムの構築を目的にしている。本稿では、聞き役対話システムの対話の流れを制御する対話制御部を自動的に構築する手法を提案した。人同士による聞き役対話を大量に収集し、それぞれの対話にユーザ満足度を付与し、このデータを

用いて POMDP で満足度の高い対話の流れを学習した。実験では、我々の提案手法である POMDP を用いたシステムと、比較用システムを実験参加者が評価した。結果として POMDP は他のシステムよりも主観評価において、有意に満足度が高いことがわかった。

今後は、今回のようなシミュレーションではなくリアルユーザに対話制御部の評価を行い、より実システムに近い状態での評価を行う。その際には、満足度の報酬を使わない EvenPOMDP だけでなく、自然性の報酬を使わない POMDP も評価する。また、対話行為ごとに適切な評価値を割り振る手法、一発話内で複数の対話行為を行う手法などの検討を行い、より柔軟に対応ができる POMDP による対話制御を構築する。その後は、ユーザの発話文理解や、システムの表層生成などを行っていく予定である。

謝辞

本研究の一部は、科研費（新学術領域）「人とロボットの共生による協創社会の創成」における計画研究「ロボットのコミュニケーション戦略の生成」（21118004）の助成を受けたものである。

参考文献

- [1] R. M. Maatman, Jonathan Gratch, and Stacy Marsella. Natural behavior of a listening agent. *Lecture Notes in Computer Science*, Vol. 3661, pp. 25–36, 2005.
- [2] Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. Analysis of listening-oriented dialogue for building listening agents. In *Proc. 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL)*, pp. 124–127. Association for Computational Linguistics, 2009.
- [3] Yasuhiro Minami, Akira Mori, Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, and Eisaku Maeda. Dialogue control algorithm for ambient intelligence based on partially observable markov decision processes. In *Proc. International Workshop on Spoken Dialogue Systems Technology (IWSDS)*, pp. 254–263, 2009.
- [4] Marilyn A. Walker, Rebecca Passonneau, and Julie E. Boland. Quantitative and qualitative evaluation of DARPA communicator spoken dialogue systems. In *In Proc. ACL*, pp. 515–522, 2001.
- [5] J.D. Williams and S. Young. Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language*, Vol. 21, No. 2, pp. 393–422, 2007.
- [6] 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, SIG-SLUD 58, pp. 61–66, 2010.
- [7] 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.