

ユーザの焦点に適応的な雑談型音声情報案内システム

吉野 幸一郎 河原 達也
京都大学 情報学研究科

概要

日々動的に更新される Web ニュース記事を対象として、音声による雑談的な情報案内を行うシステムについて述べる。ユーザがどの情報に興味があるかという焦点に着目し、ユーザとの対話を通じて漠然とした情報要求に応えることを目標とする。これを、部分観測マルコフ決定過程 (POMDP) を用いた統計的対話制御により、ユーザの状態と焦点に最適化された情報案内モジュールの選択を行うことにより実現する。

1 はじめに

これまで数多くの音声対話システムが研究開発され、一部は実世界で利用されるようになってきている。特に近年、スマートフォンなどで多様な要求に応答を行うことができるシステムも登場している。しかし、こうしたオープンドメインの対話システムにおいて、システムの応答は単純な一問一答にとどまっている。一方で、ユーザの漠然とした情報要求に対して、対象ドメインの知識を利用しながら、複数ターンにわたって対話を行うシステムも求められている。これは単純なキーワードベースの検索ではなく、観光地やレストランなどについてより詳細な情報の案内を行うものである。このようなアプリケーションは、対象とするドメインの知識を記述した文書を検索することによって実現することができる [1]。例えば、観光ガイドブックや Wikipedia 中の文書を利用して観光地のナビゲーションを行うシステム [2] や、ニュースを検索して案内を行うシステム [3, 4] が挙げられる。

近年盛んに行われている MDP や POMDP などの強化学習による対話制御では、ユーザに対する対話システムの振る舞いに対して適切な報酬を与える必要がある [5, 6, 7]。しかし、上記のような漠然とした情報要求に対しては、明確なゴールを設定して報酬を与えることが難しい。このような問題に対して、いくつかのタスク非指向型対話の研究では報酬の設計で工夫がなされている。例えば雑談モジュールの選択に強化学習を用いる枠組み [8] では、対話の終了時にユーザがその対話の満足度を評価することによって、報酬を与えている。また、聞き役対話システムの研究 [9] では、収集した対話に主観による対話満足度をアノテーシ

ョンし、システム発話とユーザ発話の組から報酬を与える識別モデルを学習している。しかし、このようなシステムの評価やコーパスのアノテーションを行うのは手間がかかり現実的ではない。

これらの先行研究に対して本研究では、ユーザが情報要求を持つタスク達成対話の性質を持つものの、それは必ずしも明確でなく、動的に変化する状況を想定する。ユーザとシステムは、システムが背後に持つ文書集合内に記述された知識について複数ターンにわたって対話を行い、その対話を通じてユーザの漠然とした情報要求にこたえる。このような想定を行った上で、対話そのものの盛り上がり (継続ターン数) や、ユーザからの積極的な対話への参与 (具体的な焦点を含む質問) などを報酬とした上で、POMDP による対話規則の獲得を行う。

2 対話による情報案内システム

自然な人対人の対話・雑談では、参加者は事前にくつつかの話題 (=トピック) を用意しており、その話題に沿って対話を進めることが多い [11]。こうした対話・雑談の形式を、情報案内と呼ぶ。例えば、まず話し手は最初的话题を聞き手に提示し、聞き手の反応をさぐる。聞き手が興味を示しながらも明確な質問をしない場合、話し手はその話題の詳細について説明する。聞き手が具体的な質問をした場合、話し手はそれに答える。一方で、聞き手が話題に興味を示さなかった場合、話し手はその話題の詳細への言及を避け、次の話題へ移行する。

ニュース案内タスク [3] ではこうした対話構造を、システムが話し手、ユーザが聞き手を担当するというより単純な状況で再現する。システムは Web ニュースから収集した話題をユーザに提供する。ユーザは興味があるニュースの情報を、対話的に得ることができる。

本稿で提案する情報案内システムの概要を図 1 に示す。システムは異なる機能を持つ 6 つのモジュールからなる。システムはユーザの発話とその述語項構造解析結果および焦点から、適切なモジュールを選択して応答を生成する。

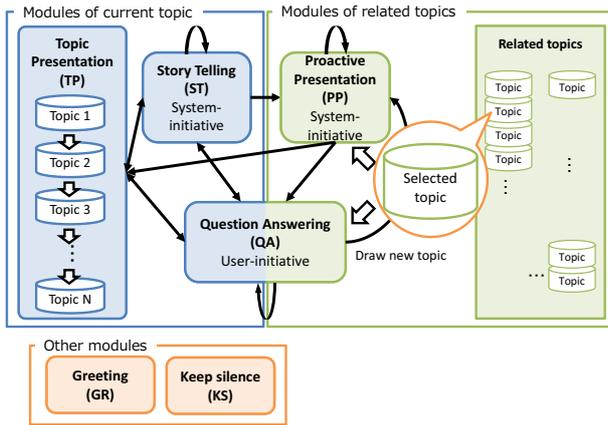


図 1: 情報案内システムの概要

2.1 システムの概要

システムはユーザの様々な要求に応えるために、6つのモジュールを持っている。Topic presentation (TP) では、用意している話題を順番に提示していく。ユーザが興味を示した場合、システムは Story telling (ST) を行う。ここでは、興味を示された話題を詳細に説明する。さらに話題への興味が継続される場合、システムは Proactive presentation (PP) によって関連する話題の案内を行う。もしユーザが具体的な質問と共に興味を示した場合、システムは Question answering (QA) によって質問応答を行う。この (QA) モジュールは、提示した話題と関連話題双方に対応する。これらのモジュール間では、図 1 に示した状態遷移が可能である。その他に、Greeting (GR) と Keep silence (KS) のモジュールが存在する。これらはいずれのモジュールからも遷移され、いずれのモジュールへも遷移可能である。KS はユーザの要求には存在せず、代わりに No request (NR) が存在する。

2.2 情報案内におけるユーザの焦点

談話構造解析においては「焦点」は「所与の時点における物体、状態、関係、発話意図などの情報への顕著な注意状態」とされている [12]。これはユーザの焦点がユーザの注意状態を示し、ユーザの発話に焦点となる対象が存在する場合、その対象に注意を向けているということになる。この定義を踏まえた上で、我々は情報案内における焦点を「ユーザの興味状態に沿った情報案内を行う上で不可欠な対象」と定義する。例えば、ユーザの質問が「鳥谷は活躍しましたか?」のような場合、「鳥谷」に焦点があるととらえる。このような焦点を、本システムにおける検索単位である述語項構造の構成要素の中で布尔値として認定する。この焦点を、ユーザ状態の推定結果とあわせて、POMDP による統計的対話制御に利用する。この焦点の推定に

は条件付き確率場 (CRF) による識別モデルを用いる。

3 情報案内のための対話制御

従来のタスク指向対話システムで用いられる対話制御は、設定されたタスクのゴールにいかにか早く到達するかということを目指したものである [6]。これと比較して、本研究で考える情報案内タスクでは常に明確な達成目標があるわけではない。そこでユーザの対話への参与度 (興味の度合い) と、ユーザの焦点 (興味の対象) を用いて対話制御を行う。本章では一般的な POMDP による対話制御を述べ、その上で提案する対話制御について述べる。

3.1 POMDP による統計的対話制御

この節では、POMDP による対話制御について述べる。まず、それぞれの表記について以下に示す。

- $s \in I_s$: ユーザ状態
ユーザがどのモジュールに対する要求を行っているかを示す。
- $a \in K$: システムの行動
システムがどのモジュールを選択するかを示す。
- $o \in I_o$: 観測状態
音声認識や意図解析の誤認識を含むユーザ状態の観測結果である。
- $b_i = P(s_i | o^{1:t})$: 信念状態
各ユーザ状態に対する確信度を確率変数で表す。
- π : 政策関数
信念状態 b を入力として行動 a を出力する。 π^* は学習によって得られた最適の政策関数である。
- r : 報酬関数
ユーザ状態 s とシステムの行動 a の組に対して報酬を与える。

POMDP による対話制御では、それまでの時系列 $1 \sim t$ の観測結果の履歴 $o^{1:t}$ に対して、最適な行動 \hat{a}^t を出力することを目的とする。

次に、信念の更新について説明する。これには観測確率と遷移確率が含まれる。時刻 t における状態 s_i に対する信念の更新式は次のように定義される。

$$b_{s'_j}^{t+1} \propto \underbrace{P(o^{t+1} | s'_j)}_{\text{Obs.}} \sum_{s_i} \underbrace{P(s'_j | s_i, \hat{a}_k)}_{\text{Trans.}} b_{s_i}^t. \quad (1)$$

Obs. は観測の尤もらしさを表現する観測確率であり、**Trans.** はユーザ状態の遷移確率である。各状態 s_i^t の信念 $b_{s_i}^t$ が更新されると、そこからシステムは最適な行動 \hat{a} を以下の政策で導く。

$$\hat{a} = \pi^* (\{b_{s_i}^t\}_{s_i \in I_s}). \quad (2)$$

表 1: 報酬

状態 s	焦点 f	行動 a					
		TP	ST	QA	PP	GR	KS
TP	0						
	1	+10	-10	-10	-10	-10	-10
ST	0						
	1	-10	+10	-10	0	-10	-10
QA	0						
	1	-10	+10	+10	-10	-10	-10
PP	0						
	1	-10	+10	+30	+10	-10	-10
GR	0						
	1	-10	-10	0	+10	-10	-10
NR	0						
	1	+10	-10	-10	-10	+10	-10
	0						
	1	-10	+10	-10	+10	-10	0

3.2 POMDP の学習

最適な政策 π^* を導くためにいくつかのアルゴリズムが提案されているが、本研究では古典的な強化学習である Q 学習 [13] を用いる。Q 学習では、時刻 t における信念 b^t に対し行動 a^t を行った場合の、将来の報酬の期待値の最大化を保証する [13]。Q 学習は以下の更新式で実現される。

$$Q(b^t, a^t) \leftarrow (1 - \varepsilon)Q(b^t, a^t) + \varepsilon[R(s^t, a^t) + \gamma \max_{a^{t+1}} Q(b^{t+1}, a^{t+1})]. \quad (3)$$

ここで ε は学習率 (ε -greedy 法) であり、 γ は将来の報酬に対する割引率である。今回は経験的に、 $\varepsilon = 0.01$ 、 $\gamma = 0.9$ を用いた。最適な政策 π^* は、Q 関数によって以下の通りとなる。

$$\pi^*(b^t) = \operatorname{argmax}_{a^t} Q(b^t, a^t). \quad (4)$$

ただし、信念 b は確率変数であることから、全ての b に対して Q 関数を計算することは困難である。そこで、Grid-based Value Iteration [14] によって、以下の関数で与えられる b について Q 関数の学習を行う。

$$b(s) = \begin{cases} \eta & \text{if } s = i \\ \frac{1-\eta}{|I_s|} & \text{if } s \neq i \end{cases} \quad (5)$$

ここで η は状態 s が i である信念の値であり、0.0 から 1.0 まで、0.1 刻みで用いる。これに加えて、意図が曖昧な状態 (全状態の確率が同じ) という状態を導入する。また、信念更新における観測確率も同様の関数で与える。

3.3 焦点を用いた対話制御

提案する対話制御では、2 種類の状態 (ユーザ意図 s 、焦点 f) を導入する。ここで観測結果 o は音声認識結果に対してロジスティック回帰により 2.1 節で述べた 6 種類の意図に分類したものの [15] である。この o

は観測不可能な状態 f と s から出力されたもので、それぞれの隠れ状態は状態遷移確率

$$P(f^{t+1}|f^t, s^t, a^t) \quad (6)$$

$$P(s^{t+1}|f^{t+1}, f^t, s^t, a^t) \quad (7)$$

によって決定する。POMDP の学習には、最適な Q 値を学習するためのユーザシミュレータが必要となるが、本研究では対話コーパス [3] から計算したこれらの確率からユーザシミュレータを構築した。

また、式 (7) は信念更新における状態遷移確率にも用いた。つまり式 (1) の信念更新は更新前の焦点 f_l と更新後の焦点 f'_m を導入することにより、

$$b_{s'_j}^{t+1} = \underbrace{P(o^{t+1}|s'_j)}_{\text{Obs.}} \sum_i \underbrace{P(s'_j|f'_m, f_l, s_i, \hat{a}_k)}_{\text{Trans.}} b_{s_i}^t \quad (8)$$

となる。ただし f は推定・学習を容易にするためにブール値とする。

ここで報酬関数の定義が重要である。各ターン終了時に与えるものを表 1 に示す。報酬関数では、適切な行動には+10、適切ではないが許容される行動に 0、不適切な行動には-10 を与えている。また、ユーザからの積極的な質問があった場合 ($s = \text{QA}, f = 1$) に適切な行動を選択できたときの報酬を+30 に設定した。この他に、対話が長く継続する方が望ましいと考えられるため、50 ターン経過時点で+100 を与えた。また、誤った行動 ($r = -10$ の行動) が 5 回行われた場合、ユーザ側から強制的に対話を終了し ($s = \text{TM}$)、-500 の報酬 (ペナルティ) を与えた。

3.4 対話制御の評価

提案した対話制御の評価を、対話コーパスから構築したユーザシミュレータによって行った。この評価では、ユーザシミュレータの出力にノイズが含まれる割合を固定し、それぞれのノイズ割合で 100,000 対話の報酬の平均を求めた。ユーザ状態を決定的に扱う MDP ($\hat{a} = \pi^*(\hat{s})$) と、確率変数として扱う POMDP の比較を図 2 に示す。また、焦点を導入する場合としない場合の比較を図 3 に示す。MDP と POMDP を比較すると、POMDP の方がノイズに頑健なことがわかる。また、焦点を導入した場合、焦点に従った適切な応答を行うことによってより高い報酬を得ていることがわかる。特に焦点を導入しない場合と比較して、音声対話システムでよく起こる 10~20% 程度のノイズ [16] に頑健 (報酬の平均で 60~80 程度の差がある) であることがわかる。

この対話制御による対話例を図 4 に示す。例では比較的低い観測確率 ($P(o|s)$) が与えられているが、適切な行動が選択されていることがわかる。また、この例のうち 4 ターン目の曖昧な入力に対し、焦点を導入

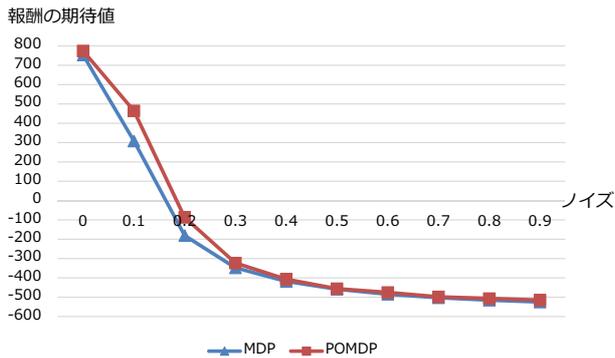


図 2: MDP と POMDP の比較

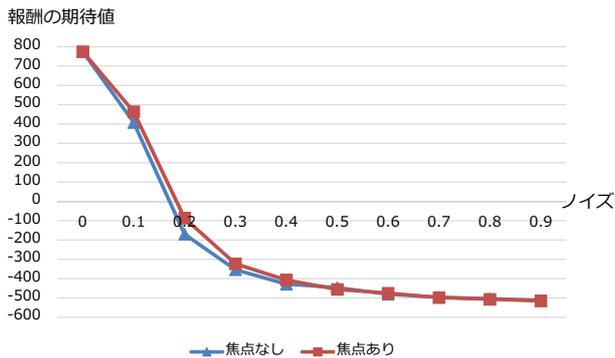


図 3: 焦点モデルの効果

した場合は $a=PP$ で応答しているのに対し、焦点を導入しないモデルでは $a=TP$ で応答していた。これはユーザの焦点がある場合、システムが自発的に情報を推薦するのが良いのではないかと考えられるためである。

4 まとめ

日々動的に更新される Web ニュース記事を対象として、ユーザの興味に応じて適応的な情報案内を行う音声対話システムを設計した。この中でユーザの焦点に対する観測モデルを導入し、これを POMDP の枠組みで変数として導入することで、適切な応答モジュール選択を行うモデルを提案した。ユーザシミュレータを用いた評価を行った結果、提案した POMDP による対話制御は MDP よりもノイズに頑健であることが確認された。また、焦点の導入によってより適切な応答モジュールの選択を行うことができることが確認された。今後、実ユーザでこのシステムを評価する予定である。また課題として、報酬関数の妥当性を検討することが挙げられる。

謝辞

この研究は JSPS 特別研究員奨励費の助成を受けた。

Sys: こんにちは。	$s=TP, P(o s)=0.8, b(s=TP)=0.800, f=0$
Usr: こんにちは。今日の野球のニュースを教えてください。	$a=TP, b(s=TP)=0.8$
Sys: マー君、ライジングカット。一緒に練習の美馬「うお！何だこれ！」	
Usr: ライジングカットってなんですか。	$s=QA, P(o s)=0.4, b(s=QA)=0.933, f=1$
Sys: ライジングカットはカットボールの軌道で深く伸びるよ。	$a=QA, b(s=QA)=0.9$
Usr: 田中はどこで練習してたの。	$s=QA, P(o s)=0.8, b(s=QA)=0.973, f=1$
Sys: 田中が20日、コボスタ宮城の室内練習場で練習をしたよ。	$a=QA, b(s=QA)=1.0$
Usr: ふーん。	$s=PP, P(o s)=0.6, b(s=PP)=0.674, f=1$
Sys: ちなみに18日の立ち投げでは、従来のものよりも沈むツーシームを練習したよ。	$a=PP, b(s=PP)=0.7$
Usr: ツーシームは沈むの？	$s=QA, P(o s)=0.6, b(s=QA)=0.836, f=1$
Sys: 18日に田中のブルペン投球を受けた長坂ブルペン捕手は「ツーシームが従来の横の変化だけでなく沈んだ」と証言していたよ。	$a=QA, b(s=QA)=0.8$
Usr: ふーん。他にニュースは？	$s=TP, P(o s)=0.8, b(s=TP)=0.982, f=0$
Sys: 藤浪もう3度目ブルペン。進化見せた。	$a=TP, b(s=TP)=1.0$
Usr: それ教えて。	$s=ST, P(o s)=0.9, b(s=ST)=0.909, f=0$
Sys: 阪神の藤浪晋太郎投手が20日、鳴尾球場の強化指定練習に参加して早くも今年3度目のブルペン入りしたよ。	$a=ST, b(s=ST)=0.9$

図 4: 対話例

参考文献

- [1] T. Kawahara. New perspectives on spoken language understanding: Does machine need to fully understand speech? In *Proc. of IEEE workshop on ASRU*, pp. 46–50, 2009.
- [2] T. Misu and T. Kawahara. Bayes risk-based dialogue management for document retrieval system with speech interface. *Speech Communication*, Vol. 52, No. 1, pp. 61–71, 2010.
- [3] 吉野幸一郎, 森信介, 河原達也. 述語項の類似度に基づく情報抽出・推薦を行う音声対話システム. *情報処理*, Vol. 52, pp. 3386–3397, 2011.
- [4] Y. Pan, H. Lee, and L. Lee. Interactive spoken document retrieval with suggested key terms ranked by a markov decision process. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20, No. 2, pp. 632–645, 2012.
- [5] E. Levin, R. Pieraccini, and W. Eckert. A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No. 1, pp. 11–23, 2000.
- [6] 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.
- [7] S. Young, M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. The hidden information state model: A practical framework for POMDP-based spoken dialogue management. *Computer Speech & Language*, Vol. 24, No. 2, pp. 150–174, 2010.
- [8] 江頭勇佑, 柴田知秀, 黒橋禎夫. 雑談対話システムにおける強化学習を用いた応答モジュールの選択. 言語処理学会 第 18 回年次大会 発表論文集, pp. 654–657. 言語処理学会, 2012.
- [9] T. Meguro, R. Higashinaka, Y. Minami, and K. Dohsaka. Controlling listening-oriented dialogue using partially observable markov decision processes. In *Proc. of COLING*, pp. 761–769. Association for Computational Linguistics, 2010.
- [10] K. Yoshino, S. Mori, and T. Kawahara. Spoken dialogue system based on information extraction using similarity of predicate argument structures. In *Proc. of SIGDIAL*, pp. 59–66, 2011.
- [11] E.A. Schegloff and H. Sacks. Opening up closings. *Semiotica*, Vol. 8, No. 4, pp. 289–327, 1973.
- [12] B.J. Grosz and C.L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, Vol. 12, No. 3, pp. 175–204, 1986.
- [13] C. Watkins and P. Dayan. Q-learning. *Machine learning*, Vol. 8, No. 3, pp. 279–292, 1992.
- [14] B. Bonet. An e-optimal grid-based algorithm for partially observable Markov decision processes. In *Proc. of ICML*, pp. 51–58, 2002.
- [15] G. Tur, U. Guz, and D. Hakkani-Tur. Model adaptation for dialog act tagging. In *Proc. of IEEE workshop on SLT*, pp. 94–97. IEEE, 2006.
- [16] 吉野幸一郎, 森信介, 河原達也. 述語項構造を介した文の選択に基づく音声対話用言語モデルの構築. 人工知能学会論文誌, Vol. 29, pp. 53–59, 2014.