

意志決定を支援する音声対話システム

翠 輝久 大竹 清敬 堀 智織 柏岡 秀紀

情報通信研究機構 音声コミュニケーション研究室

teruhisa.misu@nict.go.jp

1 はじめに

情報検索型の音声対話システムをユーザが利用する際のユーザの対話目的は、検索した情報を得ることではなく、意思決定のための手段であることがある。例えば、レストランを検索する対話システムを利用するユーザの真の目的は、レストランの価格帯などの情報を調べるのではなく、候補の中から価格を基に自分の嗜好と合致したレストランを決めることであるかも知れない。本研究では、ユーザが対象のドメインに関する知識を十分に有していないために、対話システムを利用して意思決定を行うために必要な情報を収集(システムに相談)する状況を扱う。このような状況では、ユーザはシステムがどのような情報を提供できるかを知らないだけでなく、ユーザ自身の嗜好(どのような要素を重視して意思決定を行えばいいのか)にも気づいていない可能性がある。また、このような利用シーンではシステム側も同様に、ユーザがどのような要素を重視するかに関してほとんど知識がないことが多い。そこで、システムはユーザの(潜在的な)嗜好を推定した上で、ユーザが興味を持つ情報を推薦する必要がある。ただし、その際には対話の長さとのトレードオフを考慮する必要がある。

我々は、このような相談型の対話を扱うために、部分観測マルコフ過程(POMDP)を用いたユーザの知識と嗜好を考慮する対話状態のモデルを提案してきた。試作対話システムを利用して収集した対話データを利用してユーザシミュレータを構築し、強化学習により対話戦略の最適化を行った。本稿では最適化された対話戦略を実装したシステムによる被験者実験を行い、手法の有効性を示す。

2 音声対話による意思決定支援

ユーザがシステムから提示された複数の候補の中から、1つの候補を選択する状況を考える。例えば、カーナビゲーションシステムが提示した複数のレストランの中から候補を一つ選択するような状況がこれに該当し、実世界においてもしばしば起こりうる状況である。本研究では、ユーザが自身の京都観光に対する知識が乏しい状況の下で、訪れる観光スポットを選択する状況を考える。このような相談型の音声対話シ

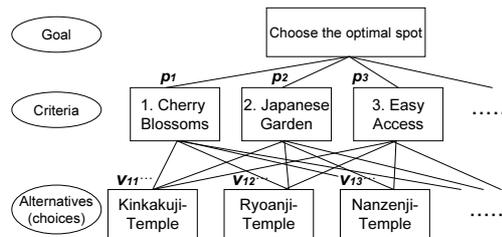


図 1: 観光案内タスク用の階層構造

テムは、意思決定支援システムの一つであると考えられる。意思決定支援タスクは、オペレーションリサーチの研究分野において多くの研究事例があり、代表的な手法として階層分析法(AHP法)[1]が提案されている。AHP法では、問題の要素を「最終目標」、「評価基準」、「代替案」の3階層に分け、ユーザの各評価基準に対する局所重み(重要度)を推定することにより最適な意思決定を行う。我々の扱う観光行為の決定支援を行う場合には、最終目的はユーザ自身の嗜好にあった観光スポットを決定することであり、代替案はシステムが紹介できる観光地のリストである。評価基準には我々が京都観光案内対話コーパスで定義している評価基準[2]を用いる。この評価基準は観光行為を確定する上で、その要因になり得るもの(たとえば、桜が有名であるとか、混雑していないなど)を指し。コーパスに付与している発話行為タグ設計の際に意識している事象の1つである。これらの要素を用いて作成した観光スポット推薦対話の階層構造の例を図1に示す。

このモデルにおいてユーザが最適な決定を行うために、評価基準に対する重み $P_{user} = (p_1, p_2, \dots, p_M)$ 、および各代替案に対する各評価基準の観点からの局所重み $V_{user} = (v_{11}, v_{12}, \dots, v_{1M}, \dots, v_{NM})$ を決定する。最適な候補の決定は、優先度 $\sum_{m=1}^M p_m v_{km}$ が最大となる代替案 k を選択することで実現される。一般的なAHP法では、評価基準や代替案に対する一対比較により上記の重みを決定がする。しかしながら、このような手法をそのまま音声対話システムに適用することは難しい。ユーザにとってシステムが提示可能な候補やドメイン知識は、対話を通じて初めて知ることができる情報である場合も多く、対話開始時点で全てが既知であることは少ない。また、システムによっては多数の候補(代替案)や評価基準を扱う場合も多

い。そのような状況下で、一対比較を行うのは非常に多くのやり取りが必要となるため、現実的ではない。そこで本研究では、システムが持つ情報をユーザに推薦しながら、ユーザの(対話開始時点でユーザ自身気づいていない潜在的なものを含む)嗜好やドメインに対する知識を推定する枠組みを考える。

3 対話システムの実装

我々が想定する音声対話システムは二つの機能を備えている。一つ目は、ユーザが要求した情報に対して、適切な情報を提供する機能である。システムはユーザが要求した場合に、当該の観光スポットに関する説明や評価基準に関する説明を提示する。二つ目は、ユーザに情報を推薦する機能である。システムはユーザから要求された情報を提供した後に、意思決定を支援するための情報を提供する。(例: システムが提供できる情報を伝達、現在の話題に対する詳細な情報を提供) 本研究で最適化するのの後者の推薦機能である。

ユーザが選択可能な京都の観光スポット 15 箇所、決定のための評価基準として 10 種類の評価基準からなるデータベースを整備した。情報検索を行う音声対話システムが扱うエントリ数と比較して候補の数が少ないが、本研究での研究対象はユーザが必須条件(例えば「京都駅付近の観光地など」)を満たす候補を比較・評価しながら決定を行うプロセスであり、このような状況において、候補数がそれほど多くないこと(15 候補)は実世界でも起こりうると思う。決定のための基準として 10 種類の評価基準からなるデータベースを整備した。前述の観光スポットに対して、スポットが各評価基準に該当するかどうかの評価 e_{nm} (“1” もしくは “0”) を人手により付与した。

ユーザに対する情報推薦の内容は、以下の 6 手法を用いて決定する。ここで、システムの情報推薦の対話行為(行動)は、発話行為 ca (=推薦手法)と意味内容 sc から構成されるものとする。発話の意味内容は観光スポットと評価基準からなり、推薦手法ごとに決められたルールに基づいて決定される。

1. 現在話題のスポットに関する情報推薦(手法 1)
直前に説明したスポットについて、詳細な説明を推薦する。具体的には、説明中のスポットに関して、知識ベース中の評価が “1” である評価基準を複数個ランダムに選んでユーザに推薦する。
2. 現在話題の評価基準に関する情報推薦(手法 2)
直前に説明した評価基準に関連した、別の観光スポットを推薦する。
3. オープンプロンプト(手法 3)
特に情報を推薦せず、オープンプロンプトを提示する。

4. 評価基準の提示 1(手法 4)
システムの現在の話題とは無関係にシステムが説明可能な評価基準を提示する。提示内容は、システムが推定するユーザの知識が低い順に選択する。
5. 評価基準の提示 2(手法 5)
手法 4 同様に、システムが説明可能な評価基準を提示する。提示内容は、システムが推定するユーザの興味が低い順に選択する。
6. ユーザが興味があると推定されるスポットの推薦(手法 6)
システムが推定するユーザの興味に最も合致する観光スポットを推薦する。この協調フィルタリングに基づいた推薦は、システムがユーザの嗜好を正しく推定できている場合に有効であると考えられるが、推定が不十分な場合には、関係のない情報を提示する可能性が高い。

4 対話戦略の最適化

4.1 ユーザモデル

最初に、ユーザのシミュレーションを行うために、知識ベクトル K_{user} 、嗜好ベクトル P_{user} 、局所重み行列 V_{user} の 3 要素からなるユーザのモデルを導入する。本研究では簡単のため、ユーザの嗜好ベクトル $P_{user} = (p_1, p_2, \dots, p_M)$ の要素は、1/0 の 2 値からなるパラメータであると仮定する。すなわち、ユーザがある評価基準 m に興味があり(もしくは潜在的に興味があり)、観光スポットを決定する際に重視する場合に p_m は “1” をとるものとする。ユーザが、(ユーザ自身も気づいていない)潜在的な嗜好を持っている状態を表現するために、ユーザの知識ベクトル $K_{user} = (k_1, k_2, \dots, k_M)$ を導入する。ユーザが、システムが評価基準 m を扱えることを知っている、もしくはシステムが手法 1、評価基準を推薦した場合にベクトルの要素 k_m は、“1” をとる。これらのベクトルを用いることで、例えば、評価基準 m が、ユーザが潜在的に興味を持っている要因であるが、ユーザはそれに気づいていないという状態は $(k_m = 0, p_m = 1)$ で表現できる。この設定は、ユーザ自身が対話開始時点から観測可能なゴール状態を持っていると仮定した従来研究(例えば [3])と対照的なものである。ユーザの評価基準 m の観点からのスポット n に対する局所重み v_{nm} は(ユーザは、システムから提示された情報のみから判断すると仮定して)、システムが推薦手法 1, 4, 5 を用いてユーザにスポットの評価を知らせた場合に “1” をとるものとする。

4.2 対話状態の記述

前節では、ユーザの対話状態の状態記述方法を定義した。しかしながら、システムはユーザの内部状

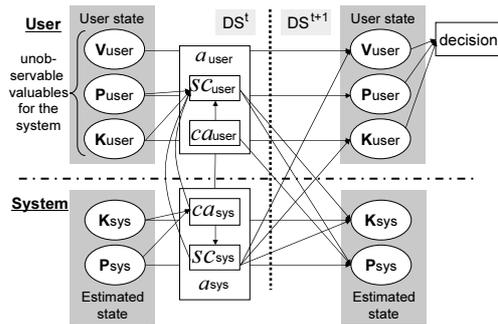


図 2: 対話モデルにおける変数の依存関係

態 ($P_{user}, K_{user}, V_{user}$) を直接観測することはできないため、ユーザとのインタラクションから推定する必要がある。そのため、このモデルは部分観測マルコフ決定過程 (POMDP) であるといえる。POMDP の状態を解決し、問題をマルコフ決定過程 (MDP) として扱うために、システムが推定するユーザの知識・嗜好の状態を表す確率分布 $K_{sys} = (Pr(k_1 = 1), Pr(k_2 = 1), \dots, Pr(k_M = 1))$ および $P_{sys} = (Pr(p_1 = 1), Pr(p_2 = 1), \dots, Pr(p_M = 1))$ を導入する。本来なら、局所重み v_{nm} に対する重みを導入することが望ましいが、本研究では v_{nm} はシステムがユーザに対して推薦した場合に“1”をとるという仮定を導入しているため、局所重みの推定は行わない。

また、ステップ (ターン) $t + 1$ における対話状態 DS^{t+1} は、直前の対話状態 DS^t とユーザ・システム間のインタラクション $I^t = (a_{sys}^t, a_{user}^t)$ のみに依存するものとする。このような近似は、対話制御を扱う多くの研究において採用され、ダイナミックベイジアンネットワークに基づく記述が行われている。本研究の対話状態における変数の依存関係を図 2 に示す。

システムが想定するユーザの嗜好 P_{sys} は、直前の対話状態 DS^t を事前分布として、ベイズ則を適用することによって更新される。ユーザの知識の推定値 k_m は、システムがユーザに評価基準を推薦した場合、もしくはユーザがシステムに当該評価基準を要求した場合に、“1”に更新される。このようにして更新された事後確率は次のインタラクション I^{t+1} を用いた対話状態更新の事前確率として用いられる。状態更新の例を図 3 に示す。

4.3 報酬関数

ユーザが選択した観光スポットが持つ属性と、ユーザの嗜好との一致率を基に報酬関数を設計する。ユーザは現在の対話状態における知識 K_{user} と局所重み V_{user} の下で、最も優先度 $\sum_m k_k \cdot p_k \cdot v_{km}$ が高いスポット k を選択するものとする。報酬 R は、ユーザが決定したスポット k が、ランダムにスポットを決定した場合と比較して、どれだけよい選択であるかに基

推定状態の事前確率:

- 知識: $K_{sys} = (0.22, 0.01, 0.02, 0.18, \dots)$
- 嗜好: $P_{sys} = (0.37, 0.19, 0.48, 0.38, \dots)$

インタラクション (状態観測):

- システムの情報推薦:
 $a_{sys} = Method1\{(Spot_5), (Det_1, Det_3, Det_4)\}$
「仁和寺 ($Spot_5$) に関しては、庭園情報 (Det_1)、世界遺産情報 (Det_3)、紅葉情報 (Det_4) が説明できます」
- ユーザ発話: $a_{user} = Accept\{(Spot_5), (Det_3)\}$
「世界遺産について (Det_3) 教えて」

推定状態の事後確率:

- 知識: $K_{sys} = (1.00, 0.01, 1.00, 1.00, \dots)$
- 嗜好: $P_{sys} = (0.26, 0.19, 0.65, 0.22, \dots)$

ユーザの知識獲得:

- 評価基準に対する知識:
 $K_{user} \leftarrow \{k_1 = 1, k_3 = 1, k_4 = 1\}$
- 局所重み: $V_{user} \leftarrow \{v_{51} = 1, v_{53} = 1, v_{54} = 1\}$

図 3: 対話状態更新の例

づいて与えられる。

$$R = \sum_{m=1}^M p_m \cdot e_{k,m} - \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M p_m \cdot e_{n,m}$$

例えば、決定したスポットが、ユーザの嗜好のうち 3 つを満たし、ランダムにスポットを決定した場合の平均の一致数が 1.3 であった場合、報酬は 1.7 となる。

4.4 シミュレーションによる評価

推薦手法の選択方法を前述の対話状態表、報酬関数、ユーザの試作対話システムログから学習したユーザシミュレータにより強化学習を用いて最適化した [4]。最適化された対話システムの典型的な挙動は、最初にユーザに評価基準に対する知識を与え (手法 4, 5)、ユーザの嗜好を推定した上で、具体的な候補を提示する (手法 6) というものである。この対話の流れは、我々が収集している対話コーパスでの人間のガイドの挙動に近い。

被験者実験に先立ち、ユーザシミュレータにより対話戦略の評価を行った。ユーザは、5, 10, 15 ターン対話を続けると仮定し、最適化した対話戦略を用いて対話ログを収集した。シミュレーション対話においてユーザは、4 つの嗜好を持ち、(=嗜好ベクトル P_{user} の 4 つの要素が“1”，残りの要素が“0”) 優先度 $\sum_m k_k \cdot p_k \cdot v_{km}$ が最大となるスポット k を選択するものとした。嗜好の選択には、被験者実験を行った後に行ったアンケートにより調べたユーザの嗜好の分布を用いた。ユーザの知識 K_{user} についても同様に、予備実験においてユーザがシステム推薦前に発話し

表 1: シミュレーションによる評価

対話戦略	報酬 ($\pm std$)		
	T = 5	T = 10	T = 15
最適化	0.43 (0.59)	0.68 (0.64)	0.84 (0.62)
B1	0.22 (0.52)	0.46 (0.63)	0.65 (0.63)
B2	0.02 (0.41)	0.13 (0.55)	0.26 (0.58)

た割合に基づいて設定した。ユーザの局所重み V_{user} は、ユーザが予備知識を持たないと仮定し、すべてを“0”に初期化した。システムのパラメータ P_{sys} および K_{sys} は試作システムで得られたユーザの統計情報を用いて初期化し、4.3節の報酬を用いて対話を評価した。さらに、試作システムの認識精度に基づき音声認識、音声理解誤りが16.7%でおこるものとし、誤りのシミュレーションを行った。ユーザとの5, 10, 15ターンの対話を行った際にシステムが得た報酬を表1に示す。

次に、学習された対話戦略を、以下の2つのベースライン手法と比較した。

1. 推薦なし (B1)

システムは要求された情報の提示のみを行い、推薦は行わない。これは、常に手法3を選択する場合と等価である。

2. ランダムに推薦 (B2)

システムは、選択可能な6手法からランダムに推薦手法を選択する。

表1に、これらのベースライン手法との比較結果を示す。最適化した対話戦略は、ベースライン手法と比較して有意に大きな報酬を得ることができた。

5 被験者実験による評価

前述の3つの対話戦略を音声対話システム利用の経験がない40名の被験者により評価した。被験者には、15個の候補スポットの中から一つを選択することのみを指示し、対話シナリオは与えていないが、スポットを決定したタイミングで決定を明示するために「(スポット名)に行きます」と発話するように指示した。

20名の被験者には最初に最適化されたシステムを用いて対話を行い、観光スポットを一つ決定してもらった。13名のユーザにはベースラインシステム1を利用してもらい、残りの7名にベースラインシステム2を利用してもらった。ユーザには実験終了後に、アンケートにより観光スポットを選択する際に重要な評価基準(10個中4個)を選択してもらった。ユーザはシステムが扱えるすべての評価基準の中から選択しているため、ここでの選択はシステムに関するすべての知識を得た上での選択であると言える。

対話戦略の比較を行う前に、ユーザの意思決定がユーザシミュレータで想定したように優先度 $\sum_m k_k \cdot p_k \cdot v_{km}$ に基づいたものであるかを評価した。システ

表 2: 被験者実験による評価

対話戦略	報酬	対話長	ドメイン外発話	受率率
最適化	0.85	10.1	18.2	52.5
B1	0.07	9.3	21.4	27.0
B2	0.09	11.8	21.1	-

ムが情報を提供した場合に、ユーザの知識 k_m 、評価 v_{nm} の値に1をとると仮定し、ユーザが決定したスポットが優先度が何番目に高いものであったかを評価した。決定されたスポットの平均順位は1.4であり、また、半数以上のユーザが最も高い優先度をもつスポットを選択していた。この結果は(ユーザが意識しているかは分からないが、)ユーザの決定が優先度に基づいているものであることを示していると言える。

次にユーザの最初の対話セッションの評価を行った。報酬、対話長、ドメイン外発話率、およびユーザの推薦受率を調べた。この結果を表2に示す。シミュレーション環境同様に、最適化された対話戦略を用いることにより、ベースライン手法と比較して高い報酬を得ることができた。平均報酬の0.85は、ベースライン手法の0.07, 0.09と比較して有意に($p < .05$)高かった。また、ユーザの推薦受率は52.2%であり、ドメイン外発話率と合わせてベースラインシステムよりも高かった。これらの結果はシステムがユーザの対話状態にあった適切な選択を行い、その結果としてユーザの良い意思決定につながったことを示していると言える。

6 おわりに

本稿では、音声対話による意思決定支援システムの被験者実験の結果について述べた。POMDPを用いた対話状態表現を用いて最適化した対話戦略とベースライン手法を比較し、シミュレーション環境による評価と同様に高い報酬が得られることを示した。今後ユーザのシステム利用履歴などを用いてユーザの初期状態の推定を行うとともに、複数のユーザタイプを用いた対話戦略の学習を行う予定である。

参考文献

- [1] T. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, 1980.
- [2] K. Ohtake, T. Misu, C. Hori, H. Kashioka, and S. Nakamura. Annotating Dialogue Acts to Construct Dialogue Systems for Consulting. In *Proc. The 7th Workshop on Asian Language Resources*, pp. 32–39, 2009.
- [3] J. Schatzmann, B. Thomson, K. Weilhammer, H. Ye, and S. Young. Agenda-based User Simulation for Bootstrapping a POMDP Dialogue System. In *Proc. HLT/NAACL*, 2007.
- [4] T. Misu, K. Sugiura, K. Ohtake, C. Hori, H. Kashioka, H. Kawai, and S. Nakamura. Dialogue Strategy Optimization to Assist User's Decision for Spoken Consulting Dialogue Systems. In *Proc. IEEE-SLT*, pp. 342–347, 2010.