

[観光情報学—スマートツーリズムに向けた研究動向—]

8 観光とチャットボット

—自動 FAQ, 推薦, 行動変容対話—

応
般

吉野幸一郎

理化学研究所 情報統合本部 ガーディアンロボットプロジェクト (GRP)



自然言語インタフェースとしての チャットボット

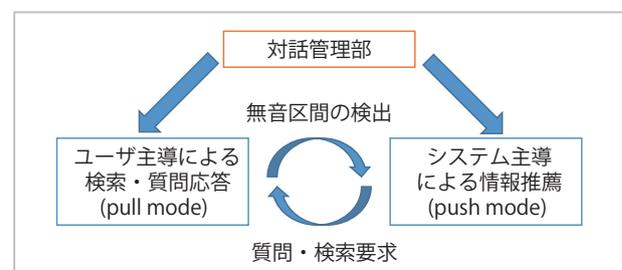
近年チャットボットはさまざまな場面で利用されるようになってきている。これは、人間の意思を伝える道具として最も大切に使いやすいものの一つが自然言語を用いたインタラクションであるためである。こうしたチャットボットを用いる場面として注目されているのが、観光や官公庁における情報案内のタスク・ドメインである。

特に近年の観光を取り巻く状況は、観光客の周りにさまざまな情報が溢れており、選択肢が多様である反面、真に必要とする情報を発見し選択することが容易ではない。Web 検索などで必要とする情報を得たい場合、必要とする情報を得るためのキーワードやハッシュタグを適切に入力する必要がある。適切なキーワードやハッシュタグを知ることは容易ではなく、またキーワード入力によって提示された情報のうち、自分に必要なものを読み取ることが必要とされる。

こうした状況を支援するために、いくつか対話的な支援を行うチャットボットのコンセプトやシステムが提案されてきた。本稿ではこれらのコンセプトやシステムを概観し、観光案内チャットボットに求められる役割や、期待される社会的役割について明らかにする。

情報コンシェルジュ

観光において、気の利いた情報を教えてくれる、必要な予約を行ってくれるといった手伝いを行ってくれる存在としては、コンシェルジュが思い浮かぶ。ホテルのコンシェルジュは顧客の質問や要求に応えるのみならず、さまざまなユーザの潜在的な要求に先回りして応えてくれる存在である。こうしたコンシェルジュの機能を実現しようとしたのが、情報爆発プロジェクトにおける「情報コンシェルジュ」¹⁾である。情報コンシェルジュは、ユーザの発話に含まれる焦点語や、サイネージを見ているユーザの視線の情報などを用いて潜在的なユーザのニーズを明らかにし、さまざまな提案を行うコンセプトとして提案された。このコンセプトでは、システムからの先回りしたプロアクティブな情報推薦 (図-1) によって、従来受動的であったチャットボットの振る舞いに能動的な要素を組み込もうとした。筆者らはこのコンセプトに従い、京都観光案内などを対象と



■ 図-1 情報コンシェルジュで提案されたプロアクティブな情報推薦

したさまざまなチャットボットを研究開発してきた(図-2)。この枠組みでは、対話中にユーザが興味を示した話題についての提示を行い、ユーザの反応、特に質問や無音状態の検出に応じて、質問応答やプロアクティブな情報推薦などさまざまな対話モジュールを呼び出す、Modular-based と呼ばれる構造を採用している。

チャットボットの形態

観光案内などを行うチャットボットを構築する上で、どのようなデバイスで何を入力として想定するかなど、いくつか形態が存在する。ここでは特に、サイネージなどの据え置き型と、スマートフォンなどのモバイル型に分類して説明する。

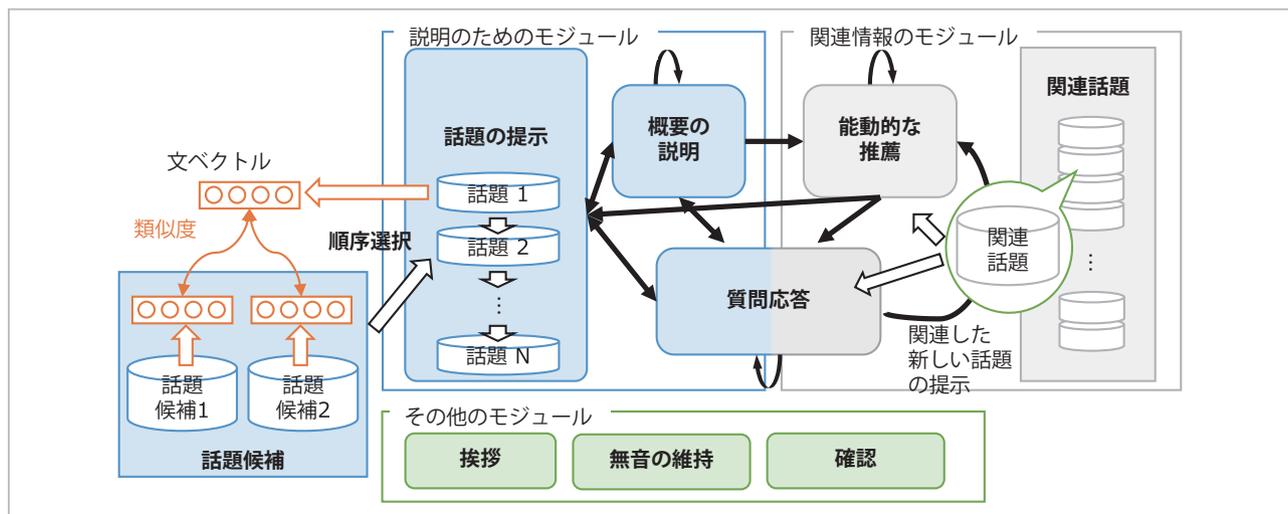
サイネージ・据え置きアシスタント

据え置き型のチャットボットは最も長く研究開発が行われてきたもので、2000年代前半から検討が行われてきた。据え置き型のチャットボットで観光案内を行う初期のもの1つとして、「たけまるくん」²⁾が挙げられる。たけまるくんは、生駒市のキャラクターの実物大筐体の実装された対話システムである。生駒市のコミュニティセンター内で実運用され、生駒の名所や天気などについて尋ねることができた。その後大型液晶パネルが安価に利用できる

ようになったなどの事情から、こうした据え置きエージェントをサイネージ上のキャラクタで置き換えるような枠組みが主流となった。代表的な例はMMDAgent³⁾を用いて開発された「メイちゃん」である。メイちゃんは名古屋工業大学に設置され、天気予報や周辺の情報についての対話を行うことができる。近年はタッチパネル式を搭載したサイネージシステムも増加しており、これに言語を用いたインタラクション機能を追加することでチャットボット相当の機能を実現する実証が行われている。

スマートフォンアプリ

特にスマートフォンが普及して以降、スマートフォンに音声インタフェースやテキストインタフェースを備えたチャットボット機能を実装することが一般的に行われるようになった。これは深層学習普及に伴う音声認識の大幅な精度向上と、スマートフォンの入力インタフェースの性能向上に由来する。スマートフォンなどのモバイル端末を利用する場合、音声認識はモバイル端末上では行わず、ユーザの発話音声、あるいは発話音声を変換した特徴量をクラウド上のサーバに送信し、サーバから音声認識結果を得て利用することが多い。こうした枠組みは単なるWeb検索の音声インタフェース付与からはじまり、「Siri」のようなスマートフォンのネイティ



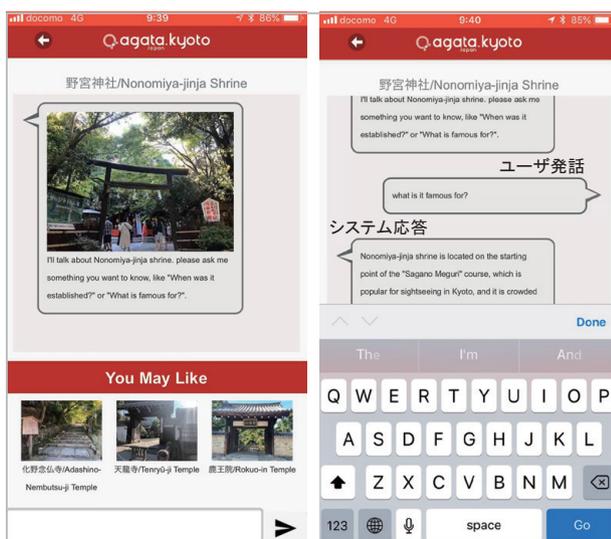
■図-2 京都観光案内チャットボットの全体像

特集
Special Feature

ブアプリ機能の呼び出しや、「しゃべってコンシェル」のような質問応答機能などが実現されてきた。さらに、観光ドメインにおいても、自治体などでもブラウザアプリやLINE 対話など、さまざまなチャットボットが実現されるに至っている。この一例として、筆者らが実際に京都の嵐山において実証実験を行った観光案内チャットボットなどがある (図-3)。

観光案内チャットボットに求められる役割

自然言語インタフェースとしてのチャットボットが普及したことにより、観光を含むさまざまな場面でチャットボットが利用されるに至っている。しかし、チャットボットが何を答えてくれるか分からず、ユーザからの利用が進まないという問題もしばしば生じている。実際のチャットボットが応答可能な範囲は限られているため、ユーザがどのような発話を行うか想定し、その発話形式にユーザを誘導することは非常に重要である。本章では、対話においてユーザ・チャットボットのどちらが主導権を握るかという観点から、観光案内チャットボットに求められる役割を整理する。



■ 図-3 京都観光案内チャットボットの対話インタフェース

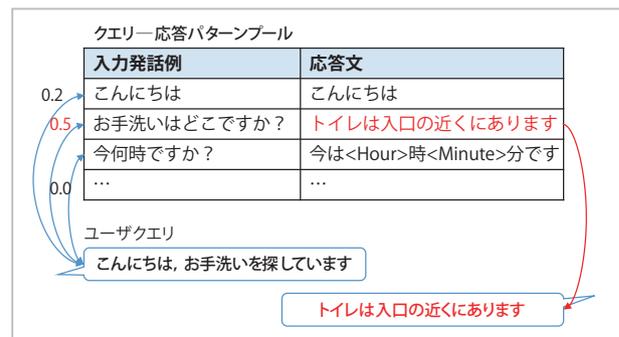
受動的応答

受動的応答とは、ユーザが何かを要求した場合にそれに対応する内容をチャットボットが答えるものである。質問応答や特定のタスク達成を目指すタスク対話などが典型的な例となるが、これらの場合、ユーザの要求は明確であり、チャットボットが行うべき行動も明らかである。チャットボットの目的は、ユーザが求める回答をいかに短時間で提示するかということになる。

これに対し、ユーザの要求が曖昧でユーザ側がクエリ・要求を明確化できない場合に、対話が止まってしまう問題がある。この問題に対応するためには、能動的応答を用いることが有効である。

パターンによる応答

従来のチャットボットの多くは受動的応答を想定しているが、ユーザの発話が多岐に渡り、オープンドメインのさまざまな対話に対応するのが難しいという問題がある。この問題に対応するために、大規模なクエリー-応答のパターンプールを用意し、この中から適切なクエリー-応答を呼び出すという response selection と呼ばれる手法が一般に用いられている (図-4)。図の例では、大量に用意されたクエリー-応答パターンプールの中から、現在のユーザクエリ「こんにちは。お手洗いを探しています」に類似する発話として「お手洗いはどこですか?」を選択し、このクエリに対応する応答である「トイレは入口の近くにありますが」という発話を応答とし



■ 図-4 Response selection

て利用している。つまり、現在のユーザ発話に最も類似するクエリを検索し、そのクエリに対応付けられた応答をチャットボットの応答として利用するという方法である。こうした枠組みは質問応答などにも用いることができ、さまざまな場面で実際のチャットボットに用いられている。また学術研究として、どのように類似するクエリを取得するか、クエリ-応答ペアになっていない候補からいかに応答候補を取得するかなどが研究されている。

このほかにも、クエリと応答の対応を学習しようとする生成的アプローチの学術研究は存在するが、生成的アプローチはチャットボットが予想しない発話を生成する可能性があるため、社会応用という観点では注意が必要である。

タスク対話（レストラン検索）

またこのほかにも、タスク対話という形が一般に用いられる。タスク対話型のチャットボットは、ユーザの要求する情報を提供するために複数回の対話を行うことを想定する。つまり、対話的に曖昧性を解消しつつユーザの検索目標を実現する。図-5の例では、ユーザの「京都のイタリアン」という発話には場所についての曖昧性がある。ここで京都全域のイタリアンを検索する、あるいは京都駅と絞り込んでしまったイタリアンを検索するという方法も考えられる。しかし、ユーザが指す「京都」の範囲を絞

り込むために問い合わせというのが、ユーザ満足度を高めるために有効な手法である。もちろん、これ以外に位置情報などのコンテキスト情報が利用可能であればこれを利用する。ただ、近年商用で利用されているチャットボットの多くではこのような対話の履歴を参照せず、ユーザの期待に沿わない検索結果が出力された場合に、ユーザがクエリを変更して入力しなおすことを期待した実装を行っているものも多い。この場合、先に述べたようにユーザが適切なクエリを入力できない可能性が高まるため、必要な追加情報をユーザに分かる形で表示することが望ましい。

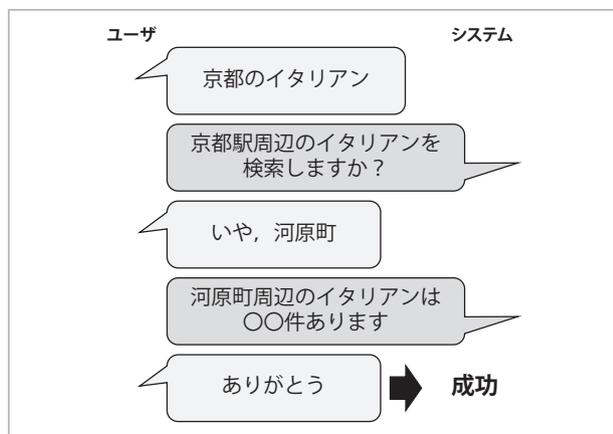
こうしたタスク対話は特に学術コミュニティでは盛んに研究が行われており、Dialogue System Technology Challenge (DSTC) や MultiWOZ などの英語での研究用オープンベンチマークが公開されている。英語においては Microsoft, Facebook, Amazon, Google などの米国 IT 大手がこうしたデータのオープンベンチマーク化を主導しており、日本語においても企業からのデータ公開が期待される。対話データをオープンベンチマーク化する場合、匿名化などの処理が必要であり、プライバシーなどにも十分配慮することが求められる。

能動的応答

ここまで議論してきた受動的応答のパターンでは、ユーザが明示的なクエリをシステムに対して与えている。しかし、実際にはユーザがクエリを思いつかない、あるいはシステムの能力を測りかねて発話を入力することが難しいといった場合も多い。こうした場合に、システムが対話の主導権を持って能動的な情報提供を行う、あるいはユーザに対してどのような情報を求めているか明示することが有効である。

システムイニシアティブ

システムが主導権を握る場合、ユーザにとっては何を話せばよいのか分かりやすい。たとえば、図-6のような例でユーザが主導権を持つ場合(ユー



■図-5 タスク対話の例

ザイニシアティブ), ユーザの発話の自由度は大きく、システムに慣れていないユーザはしばしば戸惑いを感じる。これに対してシステムが主導権を持つ場合、ユーザが何を答えればよいかは明確である。また、システム開発者側もユーザ発話を想定しやすい。しかし一方で、決まりきった対話フローで回答することを求められるのは慣れたユーザにフラストレーションを与える場合もあり、どのようなバランスでシステムが主導権を握るかについては慎重に検討する必要がある⁴⁾。

情報推薦

システムがイニシアティブを持つ場合に、ユーザに情報を求める以外にも、情報の推薦を行うことでユーザの潜在的な情報要求を喚起することができる。たとえば、ユーザが食事のジャンルを思いつかないような場合に、システムから「イタリアンはどうですか？」などと尋ねることによって、イタリアンでよいか、イタリアンよりも中華、などの追加の情報を引き出せる場合がある。こうした行動は人間のコンシェルジュも行っており、対話履歴から得られたさまざまな情報をトリガーとしてユーザが真に求める情報を探っていく。こうした対話を通じた情報の検索は、Conversational search として情報検索の分野でも検討がなされており、対話研究においても重要なトピックである。特に観光案内を対象とする場合、ユーザにとっては多様な選択肢を得ることが必要となる場合も多く、いかにユーザの潜在的欲求

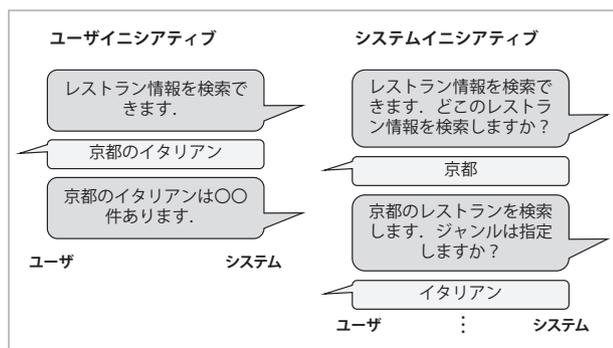
に合致した意外な情報を推薦できるかが重要となる。

対話のゴール

ここまで見てきた受動的・能動的応答を分ける上で、考慮すべき情報の一つが「対話のゴール」である。つまり、ユーザが対話の中で求めているものが何か、その求めているものがユーザの中でどの程度詳細化されているか、という点が、ユーザに対するチャットボットの振る舞いを決定する上で重要である。ユーザの中に明確なゴールがある場合、チャットボットの目的は、利用可能な情報を用いてできるだけ早くユーザの求める情報を提供することとなる。これに対して、ユーザにとってのゴールが曖昧である場合は、まずユーザのゴールを明確化することが対話の目標である。またこの場合は、ユーザの曖昧なゴールに沿う範囲で、システムがイニシアティブを発揮して情報の推薦などを行っても構わない。この点が、次の章で述べる対話を通じた行動変容に繋がってくる。

対話と行動変容

チャットボットは対話を通じてユーザに働きかけを行うことができるため、具体的な行動変容を促すことができる。たとえば、ユーザが休憩したくなるタイミングで「このカフェでは名物の○○が食べられますよ」などの情報を提示する、人流の方向を考慮しつつ全体の人流を緩和するようなルートを提示する、などの行動を行うことで、ユーザの行動変容促すことができる。こうしたユーザの行動変容を促す機械からの対話的介入については、Captology という分野で研究が行われている。特にユーザの行動やゴールの変更を促すチャットボットは説得対話システムと呼ばれ、近年盛んに研究が行われている⁵⁾。



■図-6 イニシアティブの違い

ユーザゴールとシステムゴール

チャットボットが説得対話を行う場合、行動変容の目標としてチャットボット自身(システム)のゴールを定義するということになる。このシステムゴールは、ユーザが持つ対話のゴール(ユーザゴール)としばしば衝突する場合もある。こうしたケースでどのような意思決定を行うかは、対話の方策研究として知られている。

対話の方策

チャットボットは自身の発話によってユーザの振る舞いに影響を及ぼすことができる。つまり、チャットボットの発話をシステム行動、ユーザその他から得られる情報を環境状態として、この系をマルコフ決定過程として捉えることができる。マルコフ決定過程においては、次の環境状態は現在の環境状態とそれに対応した行動によって決定される。この系において、ある特定の状態に達したとき、つまり対話においてはユーザのゴールが達成された、あるいはシステムのゴールが達成された環境状態に報酬を定義する。この将来の報酬期待値を最大化させるように、強化学習を用いて最適行動を求めることができる。つまり、最適なチャットボットの行動を対話のゴールに応じて統計的に学習することができる⁶⁾。

異なるゴールのバランス

先に述べた通り、対話におけるゴールは、特に行動変容対話においてはユーザゴールとシステムゴールが存在し、これらが衝突する場合もあり得る。衝突が生じた場合、ユーザゴールの優先度や明確度に応じてシステムゴールを変更する、システムゴールの優先度の高さによってはユーザと交渉を行うなどの対話が必要となる。こうした場合、チャットボットはユーザの意思決定を主体とする前提に立ち、対話を通じていかにユーザの同意を取り付けるかが重要な課題となる(第6章「観光ナビゲーション」)。こうしたバランスは、ユーザがチャットボットに対して求める役割や、対話内容の深刻さによっても変化するため、注意が必要である。

対話方策の統計的学習と報酬

チャットボットの振る舞いは対話方策として統計的に学習することが可能である。この学習を可能とするためには、ユーザゴールおよびシステムゴールが達成されたことを認識し、強化学習における報酬として用いなければならない。一般にタスク対話においては、ユーザからの「ありがとう」などの発話を報酬として用いる。しかし、ユーザゴールの達成・未達成は明示的に示されない場合もしばしば存在する。これに対して、ユーザの満足度や感情などを利用するアプローチも存在する。特に観光案内チャットボットにおいては、観光中のユーザ感情や満足度をさまざまな要素から推定することにより(第9章「観光客の心理状態推定」)、こうした教示を得ることができる可能性がある。ユーザの感情や満足度はレストラン案内などのタスク対話において研究用途で盛んに用いられているが、今後はより確実な生体指標などを用いることで教示とする検討も進められている。

参考文献

- 1) 河原達也, 川島宏彰, 平山高嗣, 松山隆司: 対話を通じてユーザの意図・興味を探り情報検索・提示する情報コンサルジェ, 情報処理, Vol.49, No.8, pp.912-918 (2008).
- 2) 鹿野清宏, Cincarek Tobias, 川波弘道, 西村竜一, 李 晃伸: 音声情報案内システム「たけまるくん」および「キタちゃん」の開発, 情報研報 SLP63(7), pp.33-38 (2006).
- 3) Lee, A., Oura, K. and Tokuda, K.: MMDAgent - A Fully Open-Source Toolkit for Voice Interaction Systems, In Proc. IEEE-ICASSP, 8382-8385 (2013).
- 4) 駒谷和範, 上野晋一, 河原達也, 奥乃 博: 音声対話システムにおける適応的な応答生成を行うためのユーザモデル, 信学論, Vol.J87-D-II, No.10, pp.1921-1928 (2004).
- 5) Wang, X., Shi, W., Kim, R., Oh, Y., Yang, S., Zhang, J. and Yu, Z.: Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good, In Proc. ACL (2019).
- 6) 吉野幸一郎: ニューラルネットワークによる音声対話制御, 日本音響学会誌, Vol.77, No.8 (2021).

(2021年7月30日受付)

■吉野幸一郎(正会員) koichiro.yoshino@riken.jp

理化学研究所情報統合本部ガーディアンロボットプロジェクト知識獲得・対話研究チームリーダー。音声言語処理および自然言語処理、特に音声対話、ロボット対話の研究に従事。