

観光案内への音声対話システムの活用

柏岡 秀紀 (情報通信研究機構)
水上 悅雄 (情報通信研究機構)
岩橋 直人 (情報通信研究機構)

翠 輝久 (情報通信研究機構)
杉浦 孔明 (情報通信研究機構)
堀 智織 (情報通信研究機構)

概要 本論では、情報通信研究機構で開発、公開しているスマートフォン用音声対話システム AssisTra、観光スポット推薦システム京のおすすめについて述べる。AssisTra の公開に先立ちシステムの実稼働性、有効性、ユーザビリティ等の検証、問題点の洗い出し、ユーザの行動データ収集を目的として延べ 100 名を対象に実験を行った。その概要と結果についても報告する。さらに、京のおすすめについては、システム構築のために 4000 人のアンケートを実施した。その結果を基に、ユーザが問い合わせやすい観点を見つけている。代表的な観点として「気分」「体験」「雰囲気」「特徴」に着目することにより、観光スポットを推薦するシステムである。

1. はじめに

観光立国を目指した様々な政策が立てられ、近年、ICT を観光に利用しようとする動きは、非常に活発である。このような背景の下、情報通信研究機構ユニバーサルコミュニケーション研究所音声コミュニケーション研究室では、いつでもどこでもだれとでも自然なコミュニケーションを実現するため、音声言語処理に関わる要素技術、ならびに、それら要素技術を統合したシステムの研究開発を推し進めている。音声言語処理に関わる要素技術として音声認識、音声合成、および対話制御の研究開発を進めている。さらに、これら技術を統合し、音声対話システムに関する研究開発を推し進めている。そのなかで、実用上の課題や社会からの要望を探ることを目的として、2011 年 6 月にスマートフォン用音声対話システム AssisTra (<http://mastar.jp/assisstra/>)、2012 年 3 月に同英語版 Kyo-no_hanna (http://mastar.jp/kyo-no_hanna/) を公開している。また、2011 年 10 月には、観光スポット推薦システム京のおすすめ(<http://mastar.jp/kyonoosusume/>)を公開している。これらシステムの公開により、実利用ログを利用したシステムの改良について研究すると共に、その効果を確認し、実サービスとして有効性を確認することが目的である。

本稿では、スマートフォン用音声対話システム AssisTra および観光スポット推薦システム京のおすすめの 2 つのシステムを取り上げる。AssisTra は、京都観光案内を対象に、以下の 3 つの機能から構成されている。

●はんなのガイド 京都編

「京野はんな」が京都観光ガイドとしてユーザの声を理解し、京都の観光案内をする音声対話による案内を実現した機能

●KyoTra

京都の観光名所の検索や近隣の観光地を 4 カ国語(日本語・英語・中国語・韓国語)に対応して調べることができる Web 情報を提供する機能

●TraMemo

旅の思い出として、写真の撮影や撮影した写真から自分が観光スポットを記録として保存する機能
本稿では、音声対話システムである、はんなのガイド京都編に重点をおき、その実現について述べる。また、京のおすすめは、ユーザが問い合わせやすい観点を見つけるために「気分」「体験」「雰囲気」「特徴」に着目し、観光スポットを推薦するシステムである。

2. スマートフォン用音声対話システム

本章では、スマートフォン用音声対話システムとして公開している AssisTra の機能の一つである、はんなのガイド京都編の概要と各要素技術について述べる。

2.1 公開システムの概要

AssisTra の機能の一つであるはんなのガイド京都編は、観光案内において必要な情報に容易にアクセスするための音声対話システムである。はんなのガイド京都編を開発するに当たり、単なる質問応答をする対話システムではなく、プロの観光ガイドのように前後の関連を考慮して質問応答を行うことができる対話システムを目指して研究開発を進めている。全体のシステム構成は、図 1 に示すように、大きく 5 つの要素技術から構成される。1) ユーザの発声した音声を認識しテキストとして出力する音声認識、2) 認識されたテキストの発話の内容・意図を推定する音声言語理解、3) 文脈を考慮した処理を実行し

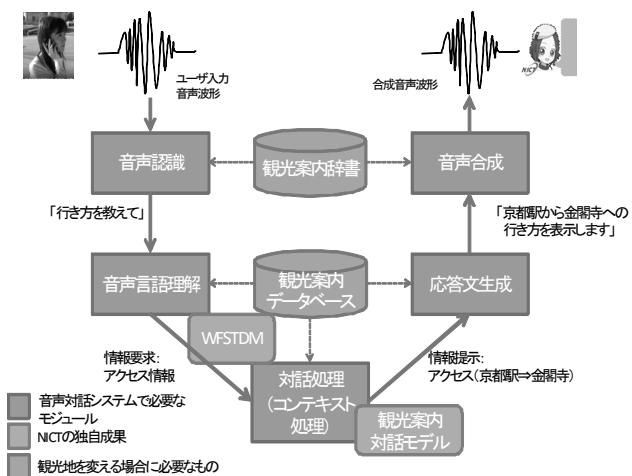


図 1 はんなのガイド 京都編

システムの行為と応答内容を確定する対話処理, 4)ユーザへの発話すべき表現を生成する応答文生成, 5)生成された表現を音声として出力する音声合成. 次節より各要素技術について、その機能と特徴について述べる。

2.2 各要素技術の概要

2.2.1 音声認識

音声認識は、隠れマルコフモデルに基づく音響モデルと Ngram による言語モデルを重み付き有限状態トランスデューサ (WFST: Weighted. Finite. State. Transducer) として表現し、結合、最適化することにより、大語彙に対応した高速/高精度な音声認識を実現している。音声認識単体としては、65万語彙を処理可能なシステム[1]となっている。本稿で記載している公開前実験の際には、語彙数5万語程度の音声認識システムを利用しておらず、大語彙への対応と共に高速化も実現している。

2.2.2 発話理解部

人間の自然発話には、発話者の個�性や状況によって様々な言い回しが存在する。たとえば、「観光スポットへのバスを利用したアクセス方法」が知りたいと考えている場合、ユーザの発話には多種多様な言い回しが存在する。「バスでのアクセス方法を教えて」「バスで行くとすれば、どうやっていけばいいですか」「最寄りのバス停からの行き方を教えて下さい」などの発話に対して（特別な状況でなければ）聞き手は同一の対応（観光スポットへのバスを利用した行き方について教えるという対応）をする。対話システムが人間同様に同一の対応をするためには、これらの表現から、同一の処理を導くように発話を理解する必要がある。実際には、各発話に対して、対話行為タグ・意味内容タグ[2]を付与することで、発話

理解を行う。そのため、収集した対話コーパスに対話行為タグ・意味内容タグを付与し、統計的な学習を行うことで発話理解を実現している。

2.2.3 対話処理

対話を実現するためには、ネットワーク上の情報、システム内部の DB に対して発話が要求する検索の実効等のアクションが必要となる。対話処理では、対話行為タグ・意味内容タグの組合せに応じたシステムアクションを準備し、システム内でそのアクションを実行し、その結果を状態として応答文生成に引き継ぐ。前項で例示した「最寄りのバス停からの行き方を教えて下さい」という表現だけでは、行き方を調べるために必須となる目的地が表現されていない。このような場合には、それまでの対話の履歴から、目的地を推定し、システムに対して目的地までのバスでの経路検索の処理を行い、次の発話文生成に処理を進める。目的地の推定ができない場合、あるいは、目的地の推定が不確かな場合には、目的地が不明である状態として、発話文生成に処理を進める。

2.2.4 応答文生成

システムアクションとして、Web 上のサービスのアプリケーションインターフェースを利用した戻り値を受け取った状態、曖昧な表現のためにユーザに問い合わせをすべきという状態など様々な状態に対して、ある一定のテンプレートに従った発話文を生成することで、最も単純な応答文生成を作成することができる。応答文生成で利用するテキストとして、プロのガイドの発話内容をもとに、桜、紅葉など様々な観点からの観光スポットの説明文を整備した。

2.2.5 音声合成

音声合成は、隠れマルコフモデルに基づく統計的手法を利用して、大量の音声データを学習に用いることで、人の発話音声に近い合成音を作成する手法[3]を採用している。学習データとして、読み上げられた音声データを利用した場合と、実際に対話により収録されたデータでは、後者の方が自然性が高いと言われている[4]。我々もこの点に注意し、学習用データを収録、利用している。

翻訳システム用のモデルの学習に利用したデータに加えて、後述の大量の観光案内対話データを利用して、観光案内用に特化したモデルを作成することで、ユーザに話しかけるような自然な合成音声を実現している。

3. プロのガイドを模倣するための対話制御

ユーザが観光する際に「どのような情報が要求され、どのように情報を提供すればよいか」という課題は、有用な対話システムを観光に活用するために解決しなければならない重要な課題である。必要とされる情報を単に提示すれば解決できるものではなく、真に必要とされる情報をインタラクションを通じて見出すことやその情報を提示する手順を実装することが課題となる。

このような課題を解決するために、プロの観光ガイドとユーザの対話を収集、分析を行い、プロの観光ガイドが行う対話の戦略をシステムに実装することとした。

以降、そのデータ収録、分析について述べ、3.2節で概略を示した対話制御、発話理解、システムの応答について示す。

3.1 プロのガイドを利用したデータ収録

京都観光案内対話コーパスは、京都観光案内のエキスパートガイド3名(男性1名、女性2名)が模擬旅行者に対して京都市内一日観光の計画立案を行う2者による対話である[5]。対面での対話、非対面での対話、ウイザードオブオズ形式での対話を収録している。1対話は約30分である。

対面での対話では、ガイドが、ガイド自身の持つ知識、準備されているガイドブック、地図、Web上の情報を利用し、旅行者に対して情報を提供、一日の旅程を作成していくものである。ガイドは、ヘッドセットマイクを使用し音声を収録している。旅行者は、スタンドマイクあるいはヘッドセットマイクのいずれかを用いて音声を収録している。旅行者は、20歳代から50歳代の114名(男性57名、女性57名)を対象として収録した。

非対面での対話では、ガイドと旅行者の間での情報の授受は、音声およびディスプレイ上の表示に限定されている。音声は、ガイド、旅行者とともに、ヘッドセットマイクを使用して収録している。非対面の対話では、対面対話の収録に参加したガイド1名(女性)に限定して収録した。旅行者は、延べ103名を対象としており、収録時間は約50時間となる。

WOZ(Wizard of Oz)形式での対話では、旅行者に情報を提示するディスプレイ上に、ガイドエージェントを表示した場合と表示しない場合の対話を収録している。旅行者は、表示/非表示のそれぞれで40名、計80名を対象としている。収録時間は、表示/非表示、各々約20時間となる。WOZ形式での収録では、システム側に、ガイ

ド、およびタイピストを用いて収録した。システムの発話は、一旦ガイドが発話したものをタイピストがテキストとして入力し、その入力テキストを音声合成したものである。定型的な応答に関しては、ショートカットキーを定義し、比較的早く応答できるように工夫している。しかし、ガイドが直接応答している対話に比べ、システムの応答に時間がかかる。

収録された対話の書き起しデータに対して、形態的なタグを始め、対話行為に対するタグ付けを行い、対話コーパスからの対話の流れをモデル化する。次節において、付与しているタグについて簡単に概要を示す。

3.2 対話行為・意味内容に関するタグ

対話行為タグの基本単位については、これまでに様々な議論がなされてきた。最も単純な単位として、文や発話といった単位が考えられる。しかしながら、人同士の対話には、必ずしも明示的な文末表現が現わされるとは限らず、境界が明確でない。また、単位を長くとると複数の対話行為が含まれやすいといった問題がある。そのため、対話中の事象を正確に捉えるためには、意味のまとまりを持つできるだけ短い単位に対してタグを付与することが望ましい。また、タグを音声対話システムの入力として使用することを想定すると、人手により認定された単位ではなく、機械的に認定可能な単位である必要がある。そこで、話し言葉の単位認定を目的に開発された単位判定プログラム[6]を用いて発話の単位を認定し、タグ付けの単位とする。

各単位に対して、発話行為を示すタグとして、MRDAタグセット[7]を拡張し、General, Response, Check, Constrain, ActionDiscussion, Others の6レイヤーと, Pause, WH の2つのサブレイヤーからなるタグセットを作成した。また、意味内容を表現するために、文節を単位とする依存構造を用いて、各文節に意味クラスを割り当てる。意味内容タグは、発話の意味を直接的に記述するものではなく、発話に含まれる単語に対してその属性を与えるものである。今回構築している観光案内システムにおいて重要となる属性の多くは、固有名ならびに、数値や日付、時刻などをあらわす固有表現である。ただし、一般的な名詞であっても、対象とするドメインによってその重要度は異なるため、固有表現のみならず、幅広く重要な表現を網羅的に収集する必要がある。

意味内容タグのための意味クラスには階層構造を持たせている。階層構造の最上位には、約40のクラス(activity, event, meal, spot, transportation, costなど)を設定している。階層構造は木構造で表現している。

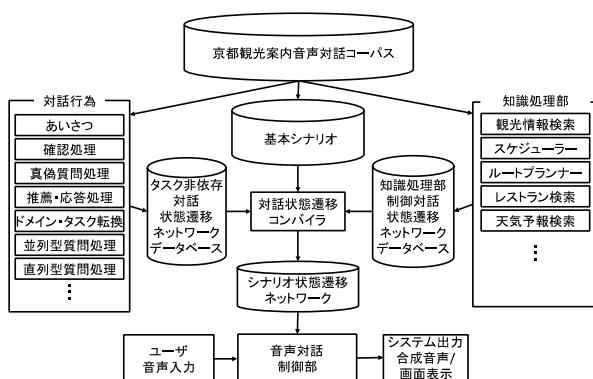


図 2 WFST による対話制御モデル

3.3 対話制御, 発話理解, システム応答

前節で示したタグを用いてコーパスより対話の流れをモデル化し、その対話の流れに従った制御をすることで、プロのガイドのような対話が可能になる。システムの対話行為タグを推定するモデル学習機構を構築し、コーパスから学習した対話制御モデルを構築した。3.1 節で述べた対話コーパスでは、オーバーラップや曖昧な表現が多く、そのままでは学習効果の評価が困難になるため、対話制御モデルの自動学習には、整備したコーパスの中から発話のオーバーラップの無い模擬対話コーパスを利用した。評価用の模擬対話において、システム側の発話の対話行為タグは、適切な対話行為タグの推定が 50% を超え、それ以外でも受理可能な対話行為タグの推定は、10%以上となった。この学習されたモデルを利用することにより、プロのガイドの対話戦略を一部システムに取り込むことができ、プロのガイドとの対話を部分的に模倣することができる。

一方で、典型的な対話のパターンについては、人が制御モデルを記述することで、対応可能となることも多い。収集したコーパスには現れていないが、人にとっては当然のような対話のパターンは、多数存在している。このようなモデルを統一的に扱うためには、コーパスから得られるモデルと人手によるモデルを統合する必用があり、WFST モデルとして図 2 に示すように両者のモデルを統合して対話の制御を行う。

また、全く同じ発話が入力された場合でも、状況や発話履歴に応じてユーザーの伝えたい意味内容が異なる場合がある。たとえば、「アクセス方法を教えて」という入力があった場合には、直前の対話のコンテキストに基づいて「どこから、どこまで、どのような交通手段で」などの情報を補完する必要がある。これらの発話に隠れた意味内容を適切に補って応答内容を決めることが必要であ

る。このような対話のコンテキストは、対話システムが利用される状況や、ユーザーがシステムを使う目的に対する依存性が高い。そこで、ユーザーの実際の利用状況に近い、前述の大規模対話データをもとに観光対話用の履歴処理モデルを作成し、対話履歴を適切に処理している。

4. 公開のための実証実験

4.1 公開前の事前実験

4.1.1 実験概要

スマートフォン用音声対話システムの実稼働性、有効性、ユーザビリティ等の検証、問題点の洗い出し、ユーザーの行動データ収集を目的とし、被験者延べ 100 名を対象に実験を行った。延べ 100 名のうち、70 名は、一回の被験者として実験に参加して頂いた。10 名は、日を経て 3 回（延べ 30 名）、実験に参加して頂いた。これは、利用者が複数回利用した際、システム利用の変化が生じるかどうかを観測するために行ったものである。実験環境としては、実際に観光しながらの利用が最も効果的であるが、通信環境、ネットワークの設定等実験環境の整備の問題から京都市内の会議室において延べ 10 日間で実験を行った。

実験に始めるに当たり、被験者には、まず、デモビデオによる説明、機能の紹介、開始終了方法の教示を行った。次に、スタッフによるサポートを受けつつ、典型的な対話となるシナリオを見ながら、対話システムを利用して頂く。その際、シナリオと多少異なる表現の入力も試すことができる。その後、実際の課題として、対話システムを利用して必要な情報を参照し、行きたい京都の観光地を 1 カ所決定することという課題を与えた。被験者に与えられた課題達成の時間は、15 分間で、1 カ所決めることができれば、再度、別の場所を決定するように、繰り返し 15 分間、課題に取り組んで頂いた。

事前に実験を行った被験者へのアンケートから、目的地や行程の決定理由、および自動音声案内システムの利用経験について図 3、図 4 に示す。

4.1.2 実験システム

実験においては、iPhone に搭載した京都観光音声対話システムで、音声認識等の処理はサーバで処理をする仕組みとなっている。また、京都の主だった観光スポット 100 カ所程度について、データベース化すると共に、京都関連の Wikipedia のページを利用して、情報を提示、Web 上にある地図サービス、レストラン検索サイトの

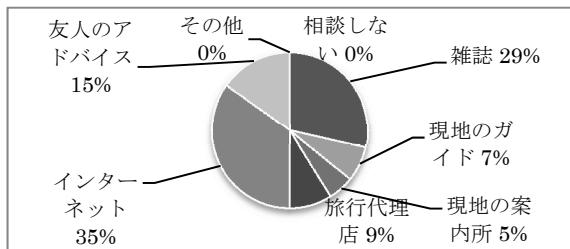


図 3 目的地や現地旅程の決定理由

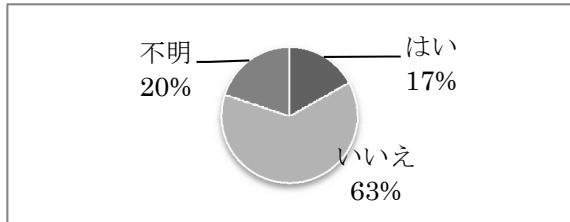


図 4 自動音声案内システムの利用経験

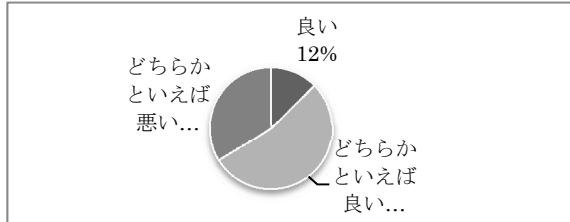


図 5 システムの全体的印象

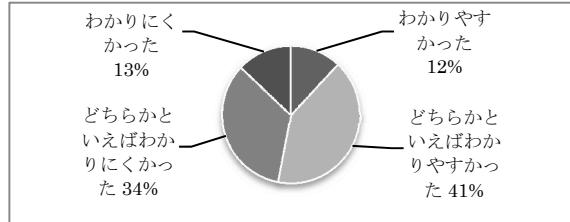


図 6 情報の提示のされ方

API を利用し、情報を提示するようにしている。

4.1.3 実験結果

延べ 100 名に対して、実験後にシステムの印象、情報提示のされ方についてアンケートを採った結果を図 5、図 6 に示す。この図からは、システムとしての印象はよく、情報提示の方法についても、からうじて半数を超える方が良いと言っている。しかし、工夫の余地は大いにある。

具体的な対話の内容については、実験実施当時、音声認識において未知語が多く、観光スポットの説明等に現れる語の入力ができないことがあった。被験者の年齢別に見た認識率・正答率を図 7 に示す。特に年齢層によって有意な差異は認められていない。

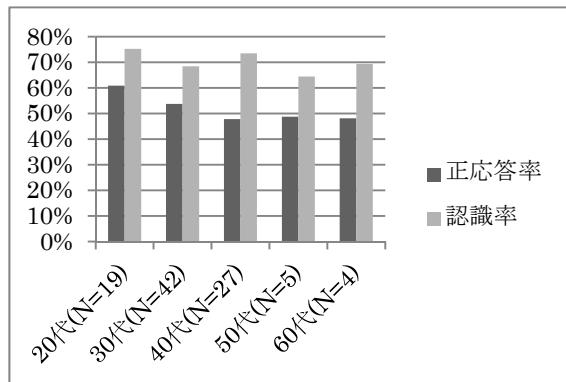


図 7 年齢別認識率・正応答率

また、システムの全体的印象と思うように話を進められた方の相関は高く、全体を通じて対話に満足した方との相関も高い。

1日おきに3回利用して頂いた被験者 10名に関して、システム全体の評価に変化は見られず、認識率、正答率も有意な差は、見られなかった。また、利用を重ねる中で、システムの持つ情報が限定的であることに気づくことで、情報の豊富さの評価が下がっている。

4.1.4 課題の抽出

実験実施時には、繰り返し課題に取り組むために、課題達成の確認により対話が終了としていた。しかし、最終確認の状態で音声認識の誤り等から確認が取れず、課題を達成したと見なされないことも生じている。被験者の実験終了後の自由記述アンケートと立ち会ったスタッフとの会話から、以下のようなシステムへの要望が多数上がっている。

●システムの持つ情報に関して：

全体的な情報量の増強、現在の情報の利用（混雑度やイベント実施など）、最新情報へのアクセス、情報の絞り込み手段のバリエーション（地域からの絞り込み、移動に対する表現）、音声と表示との一致など

●対話能力に関して：

高い認識率、理解率、曖昧な表現への対応（推量など）、高品質な合成音声、音声以外の入力デバイスなど

●個人適応に関して：

利用グループ（子供連れなどへの対応）、条件の絞り込み

●操作性に関して：

システムの状態の明瞭性、選択肢選択手法の明瞭化、地図の明瞭性向上、慣れ、戻り方、遷移等の手法の明瞭化

また、「こんにちは」で対話を開始するよう設定されて

いたが、システムに対して挨拶などの話しかけの必要性に疑問を持つ方もいた。

上記のような要望を実現するためには、言語理解部の強化、およびシステムアクションの強化、インターフェースの改善、コンテンツの強化が必用となる。また、実際の実験を連続して多人数で行うことにより、システム自体の運用の頑健性にも問題が見え、メモリ量の問題や連続動作への安定性など多様な課題が見えた。

4.2 課題への対応

音声認識に関しては、認識モジュール自体を大語彙化するために、WFSTに基づく音声認識を構築し、語彙の増強を行った。システムの持つ情報、コンテンツに関しては、既存のデータを加工し、いくつかの観点(桜、庭園、散策、紅葉など)に応じた説明を整形することにより整備を進めた。

発話理解については、実験のログ等を分析することにより順次対応を拡張している。たとえば、以下の3つの表現を考えてみる。

「有名なお寺を教えて下さい」 … (1)

「お寺で有名なところを教えて下さい」 … (2)

「仏像で有名なお寺を教えて下さい」… (3)

(1)と(2)は、ほぼ同じ意味である。しかし、(2)と(3)は、「Aで有名なBを教えて下さい」という形式で表現されており、見かけ上は同一の構造であり、同一視される。今回の実験では、(3)のような表現での入力に対して、「仏像で」という表現を考慮せずに処理されてしまい、(1)と同じ意味で処理されていた。このような表現を扱うため、副次的な属性を考慮すべき表現をコーパス内で調査し、システムアクションの生成において、考慮できるように調整を行った。また、ユーザの位置情報により左右される表現「近くのXX」等は、GPSの情報、これまでの文脈から考えられるユーザの現在地を把握するように処理を行う。文脈上の曖昧さの解消は、意味理解、および対話制御をWFSTでモデル化しており、それまでの状態遷移の経路により調整している。

4.3 公開の課題

AssisTraを公開することにより、得られたログの分析や、寄せられるコメントを見ていると、いくつかの実証のための課題が明確になって来ている。

その一つに、利用環境の問題がある。このシステムは、サーバ・クライアントの構成で実装しており、サーバ・クライアント間の通信は、電話回線での接続、あるいは無線LANでの接続を切り替えている。ユーザの移動等

により回線の状況が変化し、サーバ上の処理が済んでいる場合でも情報の転送に時間がかかり、タイムアウトが生じることがある。回線の安定性などをユーザに伝達する、クライアント側のみでタイムアウトの処理をするなどの工夫が必要となる。状況を説明するメッセージの表示をする等行っているが、さらなる工夫を行っていくつもりである。

また対話行為タグのレベルで部分的にプロのガイドの模倣を実現しているが、対話行為タグから実際の発話表現を適切な省略や補完をしながらタイミングよく生成できることも多い。収集されているログデータを分析することにより、これらの改善を進めていく予定である。

5. コンテンツへの着目「京のおすすめ」

京都やパリ、ローマに代表される観光地を訪れる場合、ガイドブック、ウェブサイト、口コミなどから情報を収集し、訪れるスポットを決める旅行者は多い。しかしながら、これらの都市には多くの観光スポットが存在するため、好みに合致するスポットを探しだす作業に必要な時間的コストは大きい。また、現状の検索技術では、「庭園がきれいで有名でないスポット」のような検索を行うことは難しい。

このような問題背景から、我々は観光スポット推薦システム「京のおすすめ」を構築した。ユーザは、気分(癒されたい、リフレッシュしたいなど)、体験したいこと、味わいたい雰囲気、観光スポットの特徴、に関連した項目をタッチパネルで選択することにより、観光スポットの推薦を手軽に受けることができる。観光スポット数は150である。



図 8 京のおすすめの動作画面 左：初期画面。
中：項目選択画面。右：スポット基本情報画面

図8左図に、システムの初期画面を示す。ユーザは「気分」「体験」「雰囲気」「スポットの特徴」の4つのカテゴリのいずれかを選択する。次に、図8中図において

カテゴリ内の項目を選択する。選択された項目により各観光スポットのスコアが計算され、画面下部に表示される。各スポットを選択すると、さらに基本情報を閲覧することができる。

カテゴリ内の項目の抽出するために、2段階の手続きを踏んだ。まず、24名の被験者に対し評価グリッド法[8]による面接を行うことにより、第1段階の項目抽出を行った。全被験者の結果を統合し、評価グリッド法において一般的な構造になるように、項目を「気分」「体験」「雰囲気」「スポットの特徴」の4カテゴリに分類した。次に、多数の被験者の意見から項目を抽出するため、ブラウザ上で動作するアンケートシステム(図9参照)を構築し、1,000名の被験者から自由記述アンケートを収集した。類語を基準として項目を統合することにより、各項目を上述の4カテゴリに割り当てた。以上の2段階の手続きにより、「世界遺産であること」「有名でない」など137の項目を得た。

このようにして得た項目を推薦に利用するために、各項目を観光スポットの属性と考え、属性のスコアを条件付き確率として定量化した。前述の手続きと同様に、ブラウザ上で動作するアンケートシステムを構築し、4,000人の被験者からアンケート結果を得た。

Q Q10	Q9の観光スポットに行って好きだった理由は、何故でしょうか。 ご自身の主觀で、例を参考にお答え下さい。 (お答えは具体的に)
〈例〉 観光スポット:(南禅寺)は、 事実や特徴等:(座禅教室がある)ので、 イメージやメリット:(新しい経験が出来た)から、 あなたの気持ち:(リフレッシュになった)と思えたので、行って好きだった。	
観光スポット	[入力欄] は、
事実や特徴等	[入力欄] ので、
イメージやメリット	[入力欄] から、
あなたの気持ち	[入力欄] 思えたので、行って好きだった。

図9 自由記述アンケート

まず、被験者に対して「行ったことがあり、好ましい」観光スポットを入力させた。次に、被験者が好むスポットについて、137の項目に関する質問を7段階(1:全く当てはまらない、2:当てはまらない、3:やや当てはまらない、4:どちらとも言えない、5:やや当てはまる、6:当てはまる、7:非常によく当てはまる)で答えさせた。以下に質問項目の例を示す。

1. 国宝級や特徴的な仏像があること・・・・・・
- 1 2 3 4 5 6 7
2. 建造物や内装が凝っていたり特徴的であるこ

と・・・1 2 3 4 5 6 7

3. 神社・仏閣であること・・・・・・・・

1 2 3 4 5 6 7

各項目を2値化し、スポットに対する各項目の条件付き確率を求めた。

図8中図では、ユーザが項目を選択する。スポットのスコアはナイーブベイズ法により計算する。すなわち、各項目が独立であるとの近似のもと、選択された条件について条件付き確率の積を求める。また、事前確率は「行ったことがあり、好ましい」と答えたユーザの割合とした。以上により、各観光スポットのスコアが計算され、画面下部に表示される。

京のおすすめは公開後、多くの方に利用されている。利用者からのフィードバックを推薦順位に反映することを検討している。ユーザが項目を選択し得られたスポットに対して、良い評価であれば、図8右図にも見られる「ここに投票！」というボタンを押下することで投票していただく。将来的には、この投票により、スポットに対する各項目の選択の条件付き確率を変更することを検討している。しかし、提示するスポットの順序に応じて、ユーザの目に触れる頻度が変わり、当初良い評価を得たものが、より多くの投票を得ることが多く、「ここに投票！」の値を反映することで推薦順位に偏りが出る問題が考えられ、投票数を利用した動的な推薦順位の変更には、まだ工夫が必要である。

また、京都以外の観光地等で、同様の推薦スポットの紹介をしたいという要望が多く寄せられている。システムの概観等については、容易に他の観光地に応用可能であるが、推薦を利用するための項目の設定、および、初期の推薦順序の設定に対して、京都で行ったアンケートと同様のアンケートを行うには、コストがかかるため、初期設定の設定方法、投票等を適切に反映する手法が、今後の課題となっている。

6. おわりに

情報通信研究機構で開発、公開しているスマートフォン用音声対話システム AssisTra、観光スポット推薦システム京のおすすめについて述べた。いつでもどこでもだれとでも自然なコミュニケーションを実現するため、音声言語処理に関わる要素技術である音声認識、音声合成、および対話制御を組み合わせて開発しており、延べ100名を対象に行った実験の概要、結果について述べた。この実験から得られた知見を生かしシステムの改良を進め、現在の公開システムとなっている。観光案内システムと

して、京都を対象として扱っているが、コンテンツ情報、語彙を組み替えることにより他の観光地に対しても容易にシステムを立ち上げることができると考えており、今後、その検証を含めて多言語、他地域でのシステムの検討を進めていきたい。

参考文献

- 1) Paul R. Dixon, Chiori Hori, Hideki Kashioka, "A Comparison of Dynamic WFST Decoding Approaches", in Proc. ICASSP, pp.4209-4212, 2012.
- 2) K. Ohtake, T. Misu, C Hori, H. Kashioka and S. Nakamura, "Dialogue acts annotation for NICT Kyoto tour dialogue corpus to construct statistical dialogue systems.", In Proc. LREC, 2010.
- 3) H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis," Speech Communication, vol.51, no.11, pp.1039–1154, Nov. 2009.
- 4) Alan W. Black, et al, "Articulatory features for expressive speech synthesis", In Proc. ICASSP, pp.4005-4008, 2012.
- 5) 大竹, 堀, 柏岡, 中村.“京都観光案内対話コーパスにおける対話行為の分析”, 言語処理学会第14回年次大会発表論文集, pp.159–162, 2008.
- 6) 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝.“日本語節境界検出プログラム CBAP の開発と評価”, 自然言語処理, Vol. 11, No. 3, pp. 39–68, 2004.
- 7) E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In Proc. 5th SIGdial Workshop on Discourse and Dialogue, pp.97–100, 2004.
- 8) 三林 紀子, 芳賀 麻薺美, 岩橋 直人, “京都観光案内対話システムのためのグレーピング評価グリッド法による選好評価構造の抽出”, 日本感性工学会第4回春季大会予稿集, 14B-01, 2009.

柏岡 秀紀（正会員）

E-mail: hideki.kashioka@nict.go.jp

1993 年大阪大学大学院基礎工学研究科博士課程修了。博士(工学)。同年 ATR 入社。1999 年奈良先端科学技術大学院大学情報科学研究科客員助教授(兼任)。2006 年情報通信研究機構。2011 年同機構室長。主に自然言語処理、音声言語処理の研究に従事。

翠 輝久（正会員）

E-mail: teruhisa.misu@nict.go.jp

2008 年京都大学大学院情報学研究科博士後期課程修了。2008 年情報通信研究機構。この間、2005~08 年日本学術振興会特別研究員(DC1)。2011~12 年南カリフォルニア大学 (USC)Institute for Creative Technologies (ICT) 客員研究員。音声言語情報処理、特に対話システムの研究に従事。博士(情報学)。2007 年度 情報処理学会山下記念研究賞受賞。2010 年度 日本音響学会栗屋潔学術奨励賞受賞。IEEE, 電子情報通信学会, 人工知能学会, 日本音響学会, 言語処理学会 各会員

水上 悅雄（非会員）

E-mail: etsuo.mizukami@nict.go.jp

1997 年神戸大学大学院自然科学研究科知能科学専攻修了。博士(理学)。2000 年学習院大学計算機センター助手, 2003 年通信総合研究所専攻研究員, 2007 年 ATR を経て, 2009 年情報通信研究機構。2012 年同主任研究員。コミュニケーションにおける調整に关心があり, 現在は対話, 話し合いの評価研究に従事。日本認知科学会, 社会言語科学会, 各会員。

杉浦 孔明（非会員）

E-mail: komei.sugiura@nict.go.jp

2007 年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。日本学術振興会特別研究員, ATR 研究員を経て, 2009 年より情報通信研究機構専攻研究員。ロボット対話, 機械学習, 推薦システムの研究に興味をもつ。計測自動制御学会, 人工知能学会, IEEE などの会員。

岩橋 直人（非会員）

E-mail: naoto.iwahashi@nict.go.jp

1985 年慶應義塾大学理工学部計測工学科卒業。1985~98 年ソニー(株)。1990~93 年 ATR に出向。1998~2004 年(株)ソニーコンピュータサイエンス研究所。2004~10 年 ATR。2008 年より, (独) 情報通信研究機構。工学博士。機械学習, 音声言語処理, 画像処理, ロボティクス, 人間-ロボットインタラクション, マルチモーダル対話システム, 言語獲得ロボットの研究に従事。ISCA, 電子情報通信学会, 日本認知科学会, 日本知能情報ファジィ学会各会員。

堀 智織（非会員）

E-mail: chiori.hori@nict.go.jp

1994 年山形大・工・電子情報卒。1997 年同大学院博士前期課程修了。1997~99 年山形大・人文・助手。2002 年東工大大学院博士(学術)。同年 NTT CS 研。2004~06 年 CMU。2007 年 ATR。2009 年情報通信研究機構。2009 年電気通信普及財団テレコムシステム技術賞受賞。

投稿受付：2012 年 7 月 29 日

採録決定：2012 年 9 月 14 日

編集担当：土井美和子(東芝)

アンケートにご協力ください。

https://www.ipsj.or.jp/15dp/enquete/enq_dp0304.html

© 2012 Information Processing Society of Japan