

3-4 対話処理システム（はんなのガイド・京のおすすめ）の開発

3-4 Development of Dialogue Systems “‘Kyo-no Hanna’ and ‘Kyo no Osusume’”

翠 輝久 水上悦雄 杉浦孔明 岩橋直人

MISU Teruhisa, MIZUKAMI Etsuo, SUGIURA Komei, and IWAHASHI Naoto

要旨

ユーザの自然な音声による情報検索要求を理解し、適切な情報を提示する音声対話システム『はんなのガイド 京都編』および、ユーザとのインタラクションを通じてユーザの潜在的な嗜好を推測し、ユーザにあった情報を推薦する推薦システム『京のおすすめ』を開発・公開した。本稿では、各システムで用いられている技術について概説する。

We developed dialogue systems of “Kyo-no Hanna” and “Kyo no Osusume”. Kyo-no Hanna interprets user's spoken queries and provide user appropriate information. Kyo no Osusume can estimates user's potential preferences and then recommends appropriate sightseeing spots that match to the user's preferences. This article abstracts several technologies used in the systems.

[キーワード]

音声対話システム, 音声言語理解, 音声認識, 推薦システム

Spoken dialog systems, Natural language understanding, Speech recognition, Recommender systems

1 まえがき

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、システムの利用者にとって自然なコミュニケーションの手法で、容易に情報システムを利用できる社会の実現を目指して、研究を進めている。我々は、人に話しかけるような、自然な音声による要求を受け付け、その意図を理解・推測することによって、適切な情報を提示する、高精度対話処理技術を研究している。また、システムとのインタラクションを通じてユーザの嗜好を適切に推定して、ユーザにあった情報を提示する技術の研究を行っている。これまでの研究成果の実証実験および実データ収集を目的として、観光案内 iPhone 用アプリ「AssisTra^{*1}」を2011年6月に（同英語版^{*2}を2012年3月に）、観光スポット推薦システム「京のおすすめ^{*3}」を2011年10月にリリースした。本稿

では、これらのアプリケーションで利用されている技術である「音声対話処理技術」および推薦システムの構築方法について概説する。

2 『はんなのガイド 京都編』

AssisTraの主要機能である『はんなのガイド 京都編』は、ユーザの自然な音声による要求を受け付け、音声と画面で、その要求に答える「音声対話システム」である。図1の例のような音声対話をすることができ、ユーザは京都の観光スポットやレストランなど観光に役立つ様々な情報を調べることができる。

一般に音声対話システムは、図2のような構

*1 <http://mastar.jp/assistra/index.html>

*2 http://mastar.jp/kyo-no_hanna/index.html

*3 <http://mastar.jp/kyonoosusume/index.html>

成をしており、大きく分けて、音声認識、音声言語理解、対話制御、応答文生成、音声合成の5つのモジュール（要素技術）で構成される。『はんなのガイド 京都編』に用いられているこれらのモジュールは、すべて当研究室で開発した。以下では、これらのモジュールについて説明する。

2.1 音声認識・音声合成・応答文生成

音声認識・音声合成は、隠れマルコフモデルに



図1 『はんなのガイド 京都編』対話例

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、自然で連続的な音声発話を認識し、人の発話音声に近い合成音を作成することができる。なお、音声認識・合成の詳細な技術については、3.2 および 3.3 の記事を参照されたい。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、より高い音声認識率と、ユーザーに話しかけるような自然な合成音声を実現している。さらに、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 音声言語理解

人間の自然発話には、ユーザー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」が知りたいというユーザーの意図を考えた場合、図3の例をはじめとして、ユーザーの発話には多種多様な言い回しが存在する。このような発話の意図を解

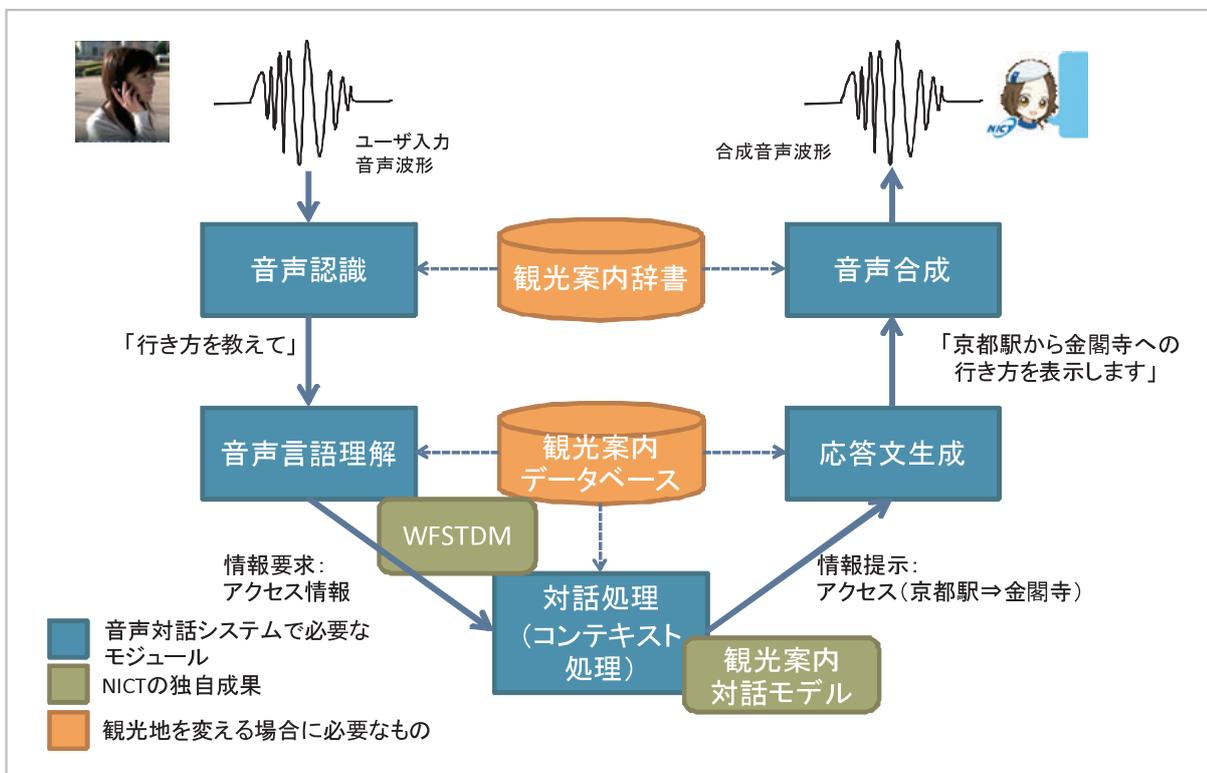


図2 音声対話システム構成図

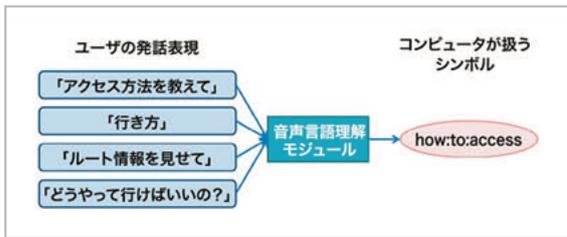


図3 言語音声理解の例

積することは人にとっては難しいことではないが、コンピュータがこれらの発話を理解するためには、これらの表現を同一のシンボル（コンピュータが処理可能な言葉）に変換する必要がある。音声言語理解がこの役割を果たす。

この機能を実現するためには、ユーザが実際に使用する表現を収集するとともに、高精度な音声言語理解アルゴリズムを研究・開発することが重要になる。会話の中で実際に利用される言い回しを収集するために、我々はプロの観光ガイドと旅行者の模擬会話を150時間300対話収録した[1]。これは、現在収集されている単一状況での音声対話データとしては世界的にも大規模なものである。さらに、プロトタイプ音声対話システムを構築して、被験者実験を行い、実際のシステム利用を想定した状況での発話表現を収集した。これらのデータをもとに、我々の研究室で独自に開発した音声言語理解・対話制御フレームワークである『重み付き有限状態トランスデューサ対話制御機構 (WFSTD: Weighted Finite-State Transducer-based Dialog Manager)』[1]を用いて、WFST表現による音声言語理解モデルを作成し、高速かつ高精度な音声言語理解を実現している。

2.3 対話制御

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発話によりユーザが期待するシステムの動作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要求を適切に補って応答内容を決めることが必要であり、対話処理部がこの役割を果たす。

このような対話履歴処理は、対話システムが利

用される状況や、ユーザ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。

2.4 はんなのガイドのまとめ

今回アプリを公開し、収集されたログデータを分析しているが、システムの応答の精度はまだ十分ではない。人間の発話や意図の種類・言い回しのバリエーションが150時間程度の学習データではカバーしきれないほど多様で複雑なものであり、コンピュータが人の意図を正確に理解するためには、より大きな対話データを収集するとともに、音声言語理解や対話履歴処理の精度の改善が必要であることが分かった。今後はシステム運用により収集した発話データを追加して各モジュールのモデルを再構築するとともに、より柔軟に発話を理解し対話を制御するアルゴリズムの研究を進めていきたい。

3 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出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。

3.1 システムの概要

図4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4中図においてカテゴリ内の評価基準（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。

3.2 評価グリッド法による評価基準の抽出

カテゴリ内の評価基準（項目）を抽出するために、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^[1]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5）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。

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段階で答えさせた。以下に質問項目の例を示す。

- 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
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4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、ある項目の値が、他の項目の値に影響しないと仮定して、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



図4 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

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。

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。
ご自身の主観で、例を参考に回答下さい。
(お答えは具付紙に)

Q9
観光スポット:(南禅寺)は、
事実や特徴等:(庭園敷地がある)ので、
イメージやメリット:(新し、遊覧が出来た)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	<input type="checkbox"/>	は、
事実や特徴等	<input type="checkbox"/>	ので、
イメージやメリット	<input type="checkbox"/>	から、
あなたの気持ち	<input type="checkbox"/>	と思えたので、行って好きだった。

図5 自由記述アンケート

4 まとめ

音声コミュニケーション研究室がリリースしたスマートフォン向けアプリケーション『ほんなのガイド』、『京のおすすめ』において利用されている技術について説明した。今後は、観光案内以外のドメイン、たとえば、ホテルやレストランの検索・推薦に適用することに加えて、対話処理および情報推薦技術の、医療や教育などの様々な分野への適用を目指していきたい。

参考文献

- 1 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.
- 2 C. Hori, K. Ohtake, T. Misu, H. Kashioka, and S. Nakamura, "Statistical Dialog Management Applied to WFST-based Dialog Systems," In Prof. ICASSP, pp. 4793-4796, 2009.
- 3 三林紀子, 芳賀麻誉美, 岩橋直人, "京都観光案内対話システムのためのグル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.

(平成 24 年 6 月 14 日 採録)



みす てるひさ
翠 輝久

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 研究員
博士 (情報学)
音声言語処理、音声対話



みずかみ えつお
水上悦雄

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
主任研究員
博士 (理学)
対話評価、コミュニケーション科学



すぎうら こうめい
杉浦孔明

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 研究員
博士 (情報学)
知能ロボティクス、機械学習



いわはし なおと
岩橋直人

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
主任研究員
博士 (工学)
知能ロボティクス、マルチモーダル対話、ヒューマンロボットインタラクション