

質問応答技術を利用したインタラクティブな観光案内システム

翠 輝久 河原 達也

京都大学 情報学研究科 知能情報学専攻

misu@ar.media.kyoto-u.ac.jp

1 はじめに

大語彙連続音声認識と情報検索技術の高度化に伴い、音声対話システムの研究対象は、関係データベースから、Web のテキストや新聞記事などの一般的な文書へと広がりつつある [1][2]。このようなシステムでは、ユーザ発話の音声認識結果と文書のマッチングが行われ、そのゆが度が高いものが提示される。

これらの文書検索システムの多くは、ユーザの検索要求にマッチした文書を提示するためにディスプレイを利用することを前提としている。しかし、携帯端末や PDA 等では、検索された文書をそのまま画面に提示することは難しい。本研究ではこのように、音声为主要なモダリティである使用環境を想定して、文書の検索・提示を行うシステムを考える。

音声による提示の場合には、ユーザの理解しやすさを考慮すると、文書を長々と読み上げるのは適切でない。ただし、文書を単純に短くすると、ユーザが知りたい重要な箇所を説明できない可能性がある。また、博物館や観光地などで利用される音声ガイドでは、たとえユーザが聞き逃しをした場合でも、その箇所を直接聞き直すことはできない。そこで、本研究では質問応答技術を利用して、文書中のあらゆる箇所に対してランダムアクセスを可能とすることにより、インタラクティブな情報提示を行うことを目指す。

しかし、このような状況下において質問応答を行うにはいくつかの問題がある。その一つに、対話コンテキストの管理が挙げられる。音声入力はテキスト入力と比較して省略や照応表現が多く、現在の発話のみを利用して適切な答えを抽出することは困難なことが多い。たとえば「いつ建てられましたか?」という発話は、この発話のみでは意味をなさない。TREC QA Track[3] などの質問応答タスクは、基本的に一問一答形式を想定してきた。NTCIR[4] など、連続した質問を扱ったものがあるが、これらは基本的に人手により作られた質問であり、実際の対話を収録したものではない。本研究では、対話のトピックの同定・追跡を行うシステムを作成し、実際の情報提示対話の中で入力される質問を扱う。

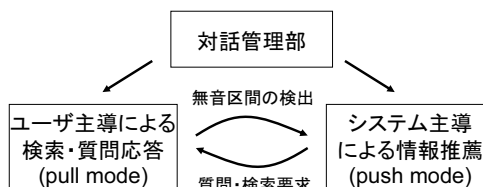


図 1: 提案システムの対話の概要

さらに、システム主導によりユーザに対して積極的に関連項目の情報推薦を行うことを考える。このような機能は、ユーザがシステムの有している情報の一覧を把握しづらい状況下において、スムーズに対話を続けるために必要であり、また、観光案内のように明確なゴールが存在せず、漠然と興味のある情報を収集する場合においても有効であると考えられる。この情報推薦をインタラクティブでユーザが受け入れやすいものにするために、質問応答技術を適用することを提案する。

以上の指針に基づいて、音声により京都の観光名所を案内するシステム「京都版ダイアログナビ」を構築し、約 3ヶ月間の運用を行った。本稿では質問応答機能を中心に、システムの評価を行った結果について報告する。

2 音声による情報検索・提示型対話システム

提案する情報提示の枠組みにおいては、ユーザ主導の検索・質問応答モードと、システム主導の情報推薦モードを用意して、ユーザの状態に応じて(対話管理部が)これらの切り替えを行う。ユーザから検索要求や質問があった場合に、検索・質問応答モードに遷移し、発話の意図にふさわしい応答を生成する。また、システム発話後から一定時間ユーザの発話がない場合には、システム主導の情報推薦モードに遷移して、現在のトピックに関連した情報を提示する。この概要を図 1 に示す。

本研究では、システムが検索・提示する文書として、Wikipedia の京都に関する文書と、京都市産業観光局が提供している京都情報データベースを使用する。そ

表 1: システムが利用する知識ベース

知識ベースの種類	件数	見出し(節)数	単語数
Wikipedia の 京都に関する文書	269	678	約 15 万
京都市観光局・ 京都情報データベース	541	541	約 7 万
合計	810	1219	約 22 万

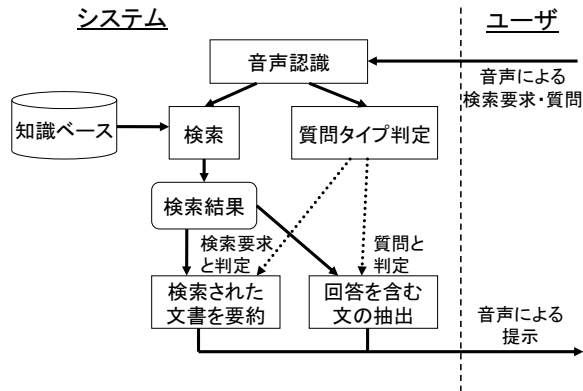


図 2: 情報検索・質問応答処理の概要

それぞれの文書の概要を表 1 に示す。

3 ユーザ主導による情報検索・質問応答

本研究ではユーザの発話を大きく二つのカテゴリに分類して扱う。一つは検索要求であり、たとえば「金閣寺について教えてください」といった発話にこれに該当する。このような発話に対しては、知識ベース中の文書の見出し(節)を単位として検索し、結果をユーザに提示する。もう一つは質問であり、「建てられたのはいつですか」といった発話にこれに該当する。このような質問に対しては、システムは知識ベースから回答を含む文を抽出してユーザに提示する。以上の手順を図 2 に示す。

3.1 トピック同定に基づく文脈解析

対話システムにおいて意味のある検索クエリを作成するために、コンテキスト情報の補完は重要である。質問文に対して照応解析を行う方法 [5] も提案されているが、これは容易ではなく、十分な精度は得られていない。そのため、本研究ではユーザの過去の発話を用いて検索に必要なキーワードを補うことを考える。しかし、単純に全てのユーザ発話を利用すると、対話中にトピックが変化した場合に、不要なキーワードを

付加してしまう可能性がある。そこで、対話のトピックを同定することで、補うべきコンテキスト長 (= 利用する過去の発話数) を決定する。

トピックの境界を検出するために、質問文の類似度を利用する手法 [6] も提案されているが、音声入力は質問文が短くなりがちであるため、このような手法を適用することは困難である。そこで本研究ではトピックとして、文書知識ベースにおいて一般的に利用可能なメタデータであるタイトル情報を利用する。すなわち、応答を生成する基となった文書を記憶・追跡することにより文脈解析を実現する。

3.2 知識ベースの検索

ユーザ発話と知識ベース中の文書との類似度を計算するために、単語ベースのベクトル空間モデルを採用する。具体的には、知識ベースの文書の節を単位として、知識ベース中の名詞に対して、タイトルに重み付けをした出現回数に基づく文書ベクトルを作成する。この文書ベクトルと、ユーザ発話の音声認識結果 (3-best 候補)・現在のトピックに関するユーザ発話 (コンテキスト情報) を音声認識の信頼度で重み付けしてベクトルに変換したものとの内積により類似度を計算する。

ユーザ発話を検索要求である場合には、検索された文書の中から、類似度が最大の見出し(節)を提示する。ユーザの理解のしやすさを考慮して、文書中の文の出現位置と文間のつながりを手がかりに重要文抽出による要約を行い、その結果を提示する。

3.3 ユーザ発話の質問タイプ判定

ユーザ発話の質問タイプの判定には、人手によるヒューリスティックなルールを用いる。たとえば、認識結果中に「誰ですか」という表現が含まれる場合には、人名をたずねる質問であると判定する。このように用意したルールにより、6種類の質問タイプに対応する。また、「なぜですか」といったシステムが回答できない質問に対しても判定ルールを用意し、ユーザがこのようなタイプの質問をした場合にも、システムが回答できないことをユーザに伝えて対処する。用意したルールのいずれにも該当しない場合には、発話は検索要求であると判定する。

3.4 質問に対する回答の生成

本研究では、テキストベースの質問応答システムで利用される一般的な手法を実装した。具体的には検索された文書の中に含まれる質問タイプに対応する固有表現 (NE) ごとに以下の特徴量を用いて回答抽出ス

コアを計算して、それが最大となる NE を質問への回答とする。

- ユーザ発話と文書間の類似度 (3.2 節)
- NE を含む文にユーザの発話した名詞が含まれる個数
- NE を含む文節に係る文節, NE を含む文節に係る文節にユーザの発話した名詞が含まれる個数

なお「創立」と「創建」のような表現の揺れの問題に対処するために、国語辞典から作成した同義語表現リストにより語彙拡張を行っている。

このようにして得られた回答をユーザに提示する際に、テキストベースのシステムのように、回答のみを答えることも考えられるが、対話の継続性などを考慮して回答を含む文全体を読み上げる。

4 システム主導の情報推薦

インタラクティブな情報推薦を行うために、システム主導の質問を生成することを提案する。質問集合は、知識ベースの文書から半自動的に生成する。また、これに加えて従来の情報推薦手法として、文書構造を利用した情報推薦と文書間類似度による情報推薦を考える。

4.1 システム主導の質問 (手法 1)

本システムではユーザの検索要求にマッチした文書を要約して提示しているが、要約の際に重要文として選ばれなかった箇所にも、有用な情報が含まれている可能性がある。このような部分をユーザに提示する方法として、システム主導でユーザに質問をすることを考える。

具体的な処理としては、事前に質問応答 (3.4 節) の回答候補となる固有表現 (NE) を含む文を対象に、以下の手順で質問文に変換しておく。

1. $tf * idf$ スコアにより、提示の対象となる NE を抽出する。
2. 文に含まれる NE を対応する疑問詞に置き換える。
3. 構文解析の情報を利用して、NE を含む文節に係る文節, NE が係る動詞を含む文節と、その動詞に係る文節以外の文節を削除する。
4. 文末表現を質問形にする。

この手法により知識ベース中の文を質問に変換した例を図 3 に示す。このようにして生成した質問から不自然なものを人手で修正・除去し、合計 495 個の質問を用意した。この中から、現在のトピックに対応する質問をユーザへの推薦の候補とする。

原文: イギリスのエリザベス女王が石庭を絶賛したことが海外のマスコミでも報道されて、...
↓ (対象の名詞を対応する疑問詞に置き換え)
- イギリスのエリザベス女王が何を絶賛したことが海外のマスコミでも報道されて、...
↓ (対象の名詞に係る/係られる文節以外を削除)
- イギリスのエリザベス女王が何を絶賛した
↓ (疑問文に変換)
質問: イギリスのエリザベス女王が何を絶賛したか知っていますか?

図 3: システム主導の質問の生成例

U1: 銀閣寺について教えてください。
S1: 銀閣寺は京都市左京区にある東山文化を代表する臨濟宗相国寺派の寺院です。通称は銀閣寺、山号は東山です。
U2: いつ建てられましたか?
S2: 銀閣寺の建立は 1482 年です。
S3: では、銀閣寺の創立者は誰か知っていますか?
U3: わかりません。
S4: 銀閣寺の創立者は室町幕府 8 代将軍の足利義政、開山は夢窓疎石とされています。
U4: どうやって行けばいいですか?
.....

図 4: システムとの対話例

4.2 文書の階層構造・文書間類似度を利用した情報推薦

システム主導の質問に加えて、一般的に用いられる情報推薦手法として、以下の 2 つの手法を実装した。

- 文書の階層構造を利用した情報推薦 (手法 2)
一般的に文書は、節などの見出しを用いて階層的に記述されていることが多い。このような文書の階層構造を利用して現在提示中の文書の未提示の節を推薦する。
- 文書間類似度を利用した情報推薦 (手法 3)
現在提示中の文書と類似度が高い文書を推薦する。

5 京都大学博物館でのシステム運用

これらの対話戦略を京都大学総合博物館企画展「コンピュータに感覚を」の展示システム「京都版ダイアログナビ」として実装を行った。システムによる対話例を図 4 に示す。

ユーザは子供から高齢者までの幅広い年齢層にわたり、大半は音声認識システムの利用経験がない。また、本システムの利用に際して、使い方などの特別な教示は与えていない。3ヶ月間の運用により 2,500 対話を収集した。この中から前半部分の対話を書き起こして、評価を行った。なお、ドメイン内の発話に対する音声認識精度 (単語認識精度) は 70.6%であった。

5.1 質問応答の性能評価

まず最初に、366回の質問により質問応答の成功率の評価を行った¹。知識ベース中にユーザの質問の回答となる箇所が存在する場合には、該当箇所を含む文を提示できた場合に質問応答成功とした。また、回答箇所が存在しない場合には、質問の回答がないことをユーザに伝えられた場合に質問応答成功とした。評価の結果、質問応答の成功率は、60.7%（回答が存在する場合で62.9%、存在しない場合で47.2%）であった。

また、音声認識結果の3-best候補をマージして用いることの効果を調べた。比較対象として、1-best候補のみを用いる場合（baseline）、3-bestの中から最適な候補を選択する場合（reference）との比較を行った。この結果を表2に示す。1-bestのみからクエリを作成する場合と比較して、3-best候補をマージしてクエリを作成することにより成功率の改善が見られた。ただし、3-best候補で別々に検索を行い、最適なものを選択することで、さらに高い成功率が得られることも示された。この成功率は、著者らが文献[7]で提案した検索後の確認を生成することによって得られると考えられる。

さらに、検索において利用するコンテキスト長（＝以前のユーザの発話数）の評価を行った。この結果を表3に示す。コンテキストの付与を行わない場合には、成功率が著しく低下している。一方、ユーザの全発話を利用した場合にも、成功率が低下しているため、必要以上の長さを付与することが成功率に悪影響を与えることが示された。従って、トピック同定に基づくコンテキスト付与の有効性が示された。

5.2 システム主導の推薦の評価

提案するシステム主導の質問の効果を確かめるために、推薦内容を利用可能な候補からランダムに選択し、ユーザに提示した。システムが生成した情報推薦は427対話中で319回であった。ユーザがシステムの提案に沿って情報提供を受けた場合に²、推薦内容を受理したものと評価した。推薦手法ごとの受理率を表4に示す。システム主導による質問（手法1）の受理率が他の手法と比較して高い。これは、手法1がユーザに回答を答えさせるという点でインタラクティブ性が高く、ユーザの興味を引くものであったためと考えられる。

¹質問応答はMRR(mean reciprocal rank)により評価が行われることが多いが、音声で複数候補を提示することが困難なため、単純な第1候補の成功率により評価を行った。

²人手により判断

表 2: N-best 候補利用と質問応答成功率の関係

N-best 候補の利用	成功率 (%)
3-best 候補のマージ (proposed)	60.7 (62.9, 47.2)
1-best 候補のみ (baseline)	57.9 (61.0, 39.6)
最適候補の選択 (reference)	63.1 (65.8, 47.2)

表 3: コンテキスト利用と質問応答成功率の関係

コンテキスト利用	成功率 (%)
現在のトピック (proposed)	60.7 (62.9, 47.2)
コンテキストなし	36.9 (30.4, 75.5)
直前の1発話のみ	54.6 (54.3, 56.6)
ユーザの全発話	55.5 (56.5, 49.1)

表 4: システム主導の情報推薦の評価

推薦手法	受理率 (%)
質問 (Proposed method 1)	74.7
文書構造 (Method 2)	51.1
文書間類似度 (Method 3)	30.8

6 おわりに

本稿では、音声为主要なモダリティである環境において、インタラクティブに文書の検索・提示を行う音声対話システムの提案を行った。観光案内システム「京都版ダイアログナビ」の構築・運用を行い、質問応答技術を利用した手法が有効に機能していることを確認した。

参考文献

- [1] A. Fujii and K. Itou. Building a test collection for speech-driven Web retrieval. In *Proc. Eurospeech*, 2003.
- [2] 西崎博光, 中川聖一. 音声認識誤りと未知語に頑健な音声文書検索手法. 電子情報通信学会論文誌, Vol. J86-DII, No. 10, pp. 1369–1381, 2003.
- [3] NIST and DARPA. The twelfth Text REtrieval Conference (TREC 2003). In *NIST Special Publication SP 500–255*, 2003.
- [4] T. Kato, J. Fukumoto, and N. Kando F. Masui. Are open-domain question answering technologies useful for information access dialogues? – an empirical study and a proposal of a novel challenge. *ACM Trans. of Asian Language Information Processing*, Vol. 4, No. 3, pp. 243 – 262, 2005.
- [5] M. Matsuda and J. Fukumoto. Answering question of iad task using reference resolution of follow-up questions. *Proc. the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, pp. 414–421, 2006.
- [6] M. De Boni and S. Manandhar. Implementing clarification dialogues in open domain question answering. *Natural Language Engineering*, Vol. 11, No. 4, pp. 343–361, 2005.
- [7] 翠輝久, 駒谷和範, 清田陽司, 河原達也. 音声対話によるソフトウェアサポートタスクのための効率的な確認戦略. 電子情報通信学会論文誌, Vol. J88-DII, No. 3, pp. 499–508, 2005.