

限定されたドメインにおける 質問応答機能を備えた文書検索・提示型対話システム

翠 輝久 河原 達也

京都大学 情報学研究科 知能情報学専攻
〒 606-8501 京都市左京区吉田二本松町
e-mail: misu@ar.media.kyoto-u.ac.jp

あらまし 音声为主要なモダリティである環境において、自然言語で記述された文書を知識源として、インタラクティブにユーザに情報を提供する枠組みを提案する。これは、音声ガイドのように、システム側から一方的に情報を提示するのではなく、ユーザ・システム双方が対話の主導権をとりながら、対話的に情報検索・提示するものである。そのために、ユーザ主導の検索・質問応答 (pull) モードと、システム主導の提示 (push) モードを用意して、ユーザの状態に応じてこれらを切り替える。検索・質問応答モードでは、漠然とした検索要求のみではなく、特定の情報・事実を求める質問応答機能も実現する。また、提示モードにおいてシステム側から、ユーザにとって有用な話題を知識ベースの中から動的に選択して (質問形式を含めて) 提示する。以上の枠組みを京都の観光案内システムとして実装し、京都大学博物館の企画展示において運用を行っている。

An interactive system for document retrieval and presentation with question-answering function in restricted domain

Teruhisa Misu Tatsuya Kawahara

School of Informatics, Kyoto University, Kyoto 606-8501, Japan
e-mail: misu@ar.media.kyoto-u.ac.jp

Abstract We propose a speech-based interactive guidance system based on document retrieval and presentation. Unlike conventional audio guidance systems, we prepare two modes of information retrieval (pull mode) and presentations (push mode), and switch them according to the user's state. In the information retrieval mode, the user can ask questions about specific facts in the documents. In the presentation mode, the system actively provides information the user would be interested in. The system was implemented as a sightseeing guidance system and is running since June 2006 at the special exhibition of the Kyoto University Museum.

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。我々も音声により文書検索を行うシステムとして「音声版ダイアログナビ」を開発してきた [3]。

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等からの利用を考えると、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。

このようなシステムにおいては、文書の提示方法が問題となる。ディスプレイを用いる場合には、検索したテキストを全て表示することもできるが、音声による出力の場合には、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、ユーザが知りたい箇所を説明できない可能性がある。

そこで、システムがユーザに一方向的に情報を提示するのではなく、マッチした文書をユーザが理解しやすい分量に区切って説明したり、ユーザが具体的な項目を質問するといったインタラクティブな情報のやりとりが必要となる。本研究では、このような説明対話を実現するために、ユーザ・システムの双方が対話の主導権をとりながら、文書にアクセスする枠組みを提案する。このシステムの設計方針に基づいて、音声により京都の観光名所を案内するシステムを構築した。

2 音声による情報検索・提示型対話システム

2.1 システムの概要

音声を主なモダリティとした文書の検索・提示を行うタスクとして、博物館等の音声ガイドや車中での情報案内のような状況を想定している。これらのタスクにおいては、現状ではシステムは決められた内容を読み上げるだけであるが、ユーザの聞き逃しや、関連情報に対する要求などに対処できると望ましい。これを実現するために、本研究では、説明中

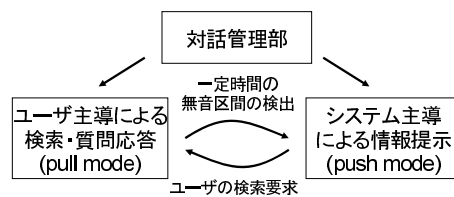


図 1: 提案システムの対話の概要

の聞き逃した項目や、説明を聞いた後の疑問などを直接システムにたずねられるようにする。これにより、文書中の知りたい箇所に対して音声によりランダムアクセスを可能とすることを目指す。

また、音声ガイドなどでは、ユーザから一定時間発話がない場合にはシステム側からユーザに積極的に関連情報の提示をしていくことが望ましい。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、トラブルシューティングなど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。

以上を実現するために、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提示モードを用意して、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枠組みを提案する。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提示モードに遷移して、システムが持つ関連情報を提示する。この概要を図 1 に示す。

検索・質問応答、情報提示に関しては、3 章、4 章において、それぞれ詳細に説明する。

2.2 システムが利用する知識ベース

本研究では、システムが検索・提示する文書として、Wikipedia¹ の京都に関する文書と、京都市産業観光局が提供する京都情報データベース² を使用する。Wikipedia からは「京都」という単語がしきい値回数以上出現する文書を抽出した。この中には、金閣寺などの名所の他に、「新選組」や「懐石料理」といった京都に関係の深い項目が含まれる。これらの文書の多くが、記事の概要の下に見出しが設けられた節による階層構造を持っている。京都情報データベースには、京都市内の寺社を中心に京都の「通り」

¹ <http://ja.wikipedia.org/>

² <http://raku.city.kyoto.jp/sight.phtml>

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

慈照寺

(概要)

慈照寺は、京都府京都市左京区にあり、東山文化を代表する臨済宗相国寺派の寺院。通称銀閣寺、山号は東山。開基は、室町幕府 8 代将軍の足利義政、...

沿革

室町幕府 8 代将軍足利義政は、1473 年、嗣子足利義尚に將軍職を譲り、1482 年から、東山の月待山麓に東山山荘の造営を始めた。この地は、...

境内

錦鏡池を中心とする池泉回遊式庭園「苔寺」の通称で知られる西芳寺庭園を模して造られたとされるが、江戸時代に改修されており、...

図 2: Wikipedia の文書の例

の名前などに関するページが存在する。本研究では、これらのページ中の 5 文程度の内容概要の項目を利用する。それぞれの文書の概要を表 1 に、Wikipedia の文書の例を図 2 に示す。

これらの文書は、Web ブラウザによりテキストの形で閲覧することを前提に作成されており、文体は書き言葉調であるため、音声合成によりそのまま読み上げるのは不自然になる。そのため、文末の助詞の変更、書き言葉特有の語彙の平易な表現への言い換えを行う [4] ことで話し言葉調に変換した。本研究では、これらの文書を構文解析器 CaboCha³ により構文解析したものを知識ベースとして使用する。

2.3 音声認識用言語モデルの構築

ユーザの発話を音声認識するための言語モデルとして、京都の観光案内タスク用に大量の Web テキストを利用して、N-gram 言語モデルを学習した。この言語モデルは、システムが利用する知識ベース (表

³ <http://chasen.org/~taku/software/cabocha/>

1) を元データとして、以下の手順で作成した [5]。まず、知識ベース中の文書中からキーワードを抽出して、検索エンジンにより関連する Web ページを収集する。次に、知識ベースと CIAIR 車内音声対話コーパスから、テキスト選択用の言語モデルを作成し、このモデルにより発話スタイルのマッチした文を Web テキストから選択する。このように選択した Web テキストと知識ベースから単語 3-gram モデルを学習した。

3 限定されたドメインにおける情報検索・質問応答

ユーザからシステムに対して検索要求・質問があった場合には、システムは検索・質問応答モードに遷移し、ユーザの要求に合った応答を生成して、音声合成により読み上げる。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「○○について教えてください」のようなある事項についての概要をたずねる発話がこのように分類される。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。もう一つはより具体的な質問であり、「○○はいつですか」といった発話がこのように該当する。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に該当する箇所を抽出する。この手順は、以下の通りであり、図 3 に処理の概要を示す。

1. 音声認識結果を用いてベクトル空間モデルにより、知識ベースを検索する。
2. ユーザ発話の質問タイプを判定する。
 - 3a. (発話が検索要求である場合)
 - ユーザの検索要求との類似度が最大の文書を要約して説明する。
 - 3b. (発話が質問である場合)
 - 検索結果の文書の中から、回答箇所を抽出してユーザに提示する。

また、それぞれの処理の詳細について、以下の節で説明する。

3.1 ベクトル空間モデルによる知識ベースの検索

本研究ではユーザの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

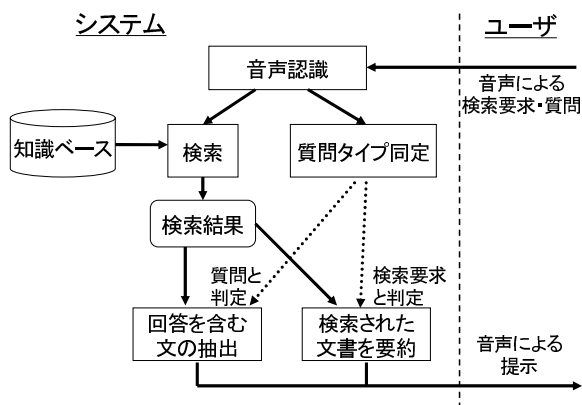


図 3: 情報検索・質問応答処理の概要

文書の節を単位として、知識ベース中の自立語に対して、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。たとえば、図2の例においては、概要、沿革、境内の説明に対して、それぞれベクトルが作られる。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) 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。なお、一連の対話の中で、コンテキストを考慮したマッチングを行うために、類似度計算には、履歴中の発話に含まれる単語も使用する。ただし、利用する履歴は現在提示中の文書に関するものに限る。

なお、ユーザの発話の音声認識誤りに対処するために、音声認識の信頼度が低い場合には、文書のタイトルや見出しを利用して「金閣寺でよろしいでしょうか」のような確認を行う。

3.2 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、音声認識結果中に「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し、「いくらですか」という表現が含まれる場合には金額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、システムが質問に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。

3.3 検索要求に対する応答生成

ユーザの発話が検索要求である場合には、検索された文書の中から、ユーザ発話との類似度が最大の見出し(節)を提示する。検索した文書をそのまま読み上げること考えられるが、ユーザの理解のしやすさを考えると、一度に説明する情報は一定の長さ以下に抑えることが望ましい。そこで、検索された節が長い場合には、一定の長さに要約をして提示する。具体的には、文書中での文の出現位置と文間のつながりを利用して重要文抽出を行い、その結果を音声合成により読み上げる。

3.4 質問に対する回答の生成

ユーザの発話が質問である場合には、知識ベースの中から回答箇所を抽出してユーザに回答する。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 (3.1 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。

4 関連度・重要度と文書構造を用いた情報提示

音声を主なモダリティとする対話システムにおいては、ユーザがシステムが有している情報の一覧を把握することは困難であり、何をたずねてよいかわかりにくい状況に陥りやすい。また、本タスクのように対話のゴールが明確でない場合には、ユーザの興味を引き出すためにも情報の推薦は有効である。そこで、ユーザから一定時間発話がない場合に、システムはこれまでの対話履歴をもとに以下の3つの手法により提示する内容を決定する。この際に、ユーザが興味を持ちそうな箇所 (=文書に特徴的な箇所) を優先的に提示するために、提示候補を $TF \cdot IDF$ ス

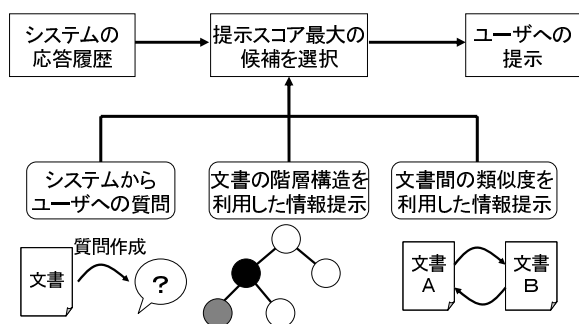


図 4: 情報提示処理の概要

コアに基づく提示スコアにより比較し、スコアが最大の候補をユーザに提示する。この処理の概要を図 4 に示す。

4.1 システムからの質問生成 (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、たとえば「より詳細な説明をしましょうか?」といったプロンプトを出すことも考えられるが、ユーザの現在の話題に対する興味を引き出すために、システムからユーザへの質問という形式を考える。

具体的な処理としては、事前に質問回答 (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. 文に含まれる NE を対応する疑問詞に置き換える。
2. 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。
3. 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 5 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、直前にユーザに提示した見出し (節) に対応する質問文をユーザへの提示の候補とする。この手法による提示スコアは、質問の回答となる語句 (疑問詞に置き換えられた NE) の TF*IDF スコアである。

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、一躍有名になりました。



生成した質問：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 5: システムからの質問の生成の例

4.2 文書の階層構造を利用した情報提示 (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。本研究で扱う Wikipedia においても、図 2 に示すように、見出し語を用いて階層的に記述されている。このような文書の階層構造を利用して情報の提示を行う。具体的には、現在提示中の文書の未提示の節を提示候補とする。たとえば、図 2 の銀閣寺の文書において、ユーザが概要部分を検索した場合には、同一文書中の節である「銀閣寺の沿革」や「銀閣寺の境内」を提示候補とする。

この手法による提示スコアには、提示対象の節に含まれる単語の TF*IDF スコア (上位 3 単語の平均値) を用いる。

4.3 文書間類似度を利用した情報提示 (手法 3)

Web 等の情報推薦においては、ユーザの表示履歴情報を利用して、ユーザの興味が近い文書が推薦することが一般的に行われている。この手法では、このような情報推薦における知見を利用して、次に提示する文書を選択する。具体的には、現在提示中の文書と類似度が高い文書を紹介する。たとえば金閣寺の説明後には足利義満、上賀茂神社の説明後には下鴨神社などが提示候補となる。この手法による提示候補の提示スコアは、文書の TF*IDF ベクトルベースのコサイン距離類似度に、提示する文書の TF*IDF スコアが高い上位 3 単語の平均値を乗じたものを用いる。

以上の対話戦略を実装したシステムによる対話の例を図 6 に示す。

U 1 : 銀閣寺について教えてください。
 S 1 : 銀閣寺でよろしいでしょうか？
 U 2 : はい。
 S 2 : 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。銀閣寺の歴史やみどころについて何か知りたいことはありますか？
 U 3 : いつ建てられましたか？
 S 3 : 銀閣寺の建立は 1482 年です。
 (一定時間の無音を検出)
 [提示内容の決定]
 候補 1 : 「銀閣寺は誰が建てたか知っていますか？」
 [手法 1](提示スコア 42.1)
 候補 2 : 「銀閣寺の庭園について説明しましょうか？」
 [手法 2](提示スコア 29.6)
 候補 3 : 「足利義政について説明しましょうか？」
 [手法 3](提示スコア 25.9)
 候補 4 : …
 [候補 1 を選択]
 S 4 : では、銀閣寺は誰が建てたか知っていますか？
 U 4 : わかりません。
 S 5 : 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。
 [次の提示内容を決定]
 S 6 : では、銀閣寺の庭園について説明しましょうか？
 …

図 6: システムとの対話例

5 京都大学博物館でのシステムの運用

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」⁴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。実装に際して、奈良先端大学院大学で開発された音声対話システム「たけまるくん」をベースに情報検索・質問応答・情報提示の機能を拡張した。また、たけまるくんを利用されているエージェントの代わりに、情報通信研究機構 (NICT) で開発されたロボット Phyno(フィノ)を利用している。このシステムでは、Phyno の顔認証機能を利用して、ユーザの名前を呼びかけるところから対話を始める (図 7)。なお、写真中のディスプレイは京都の地図や観光地の写真などを表示するもので、ユーザに説明するテキストの表示に使用されることはない。

今後 3ヶ月間の展示期間において収集される対話のログデータを分析し、システムの改善と評価を行っていく予定である。

⁴ <http://www.museum.kyoto-u.ac.jp/japanese/event/exhibition060607.html>



図 7: 京都大学博物館でのシステムの利用風景

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。ユーザが効率的に知識を得ることができるように、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提示モードを用意して、対話の状況に応じてこれらを切り替える。今後、システムを運用していくことにより、システムの改善や対話戦略の評価法の検討を行っていく予定である。

謝辞

本研究に対して多大な協力を頂いた奈良先端科学技術大学院大学の鹿野清宏先生、京都大学の黒橋慎夫先生、笹野遼平氏、及び正司哲朗氏に深く感謝します。

参考文献

- [1] A. Fujii and K. Itou. Building a test collection for speech-driven web retrieval. In *Proc. EUROSPEECH*, Vol. 2, pp. 1153–1156, 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] 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. 信学論, Vol. J88-DII, No. 3, pp. 499–508, 2005.
- [4] N. Kaji, M. Okamoto, and S. Kurohashi. Paraphrasing Predicates from Written Language to Spoken Language Using the Web. In *Proc. of Human Language Technology Conference (HLT/NAACL2004)*, 2004.
- [5] 翠輝久, 河原達也. 音声対話システムのための web テキストの選択による効率的な言語モデル構築. 日本音響学会研究発表会講演論文集, 2-11-12, 春季, 2006.