

# 音声による文書検索タスクにおける 検索要求明確化のための対話戦略

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、話し言葉の音声対話により Web のテキストや新聞記事などの一般的な文書の検索を行うシステムの研究も可能になった。NTCIR-3 の Web 検索タスクにおいても、音声入力サブタスクが提案され、その評価が行われている [1, 2]。また、我々も大規模なソフトウェアサポートマニュアルを話し言葉音声により検索するシステムとして「音声版ダイアログナビ」を構築し、試験運用を行っている [3]。

このような大規模な知識ベースを対象としたシステムにおいて、ユーザが意図する文書を正しく同定するためには、発話文中に検索に必要な情報が十分に含まれている必要がある。従来の文書検索・質問応答タスクでは、ユーザの検索要求が具体的であることを前提として、それに対する回答を提示することが目標とされているため、システムは基本的に一問一答形式である。しかしながら、話し言葉では発話が断片的になったり、検索要求が漠然なまま入力されることが多い。また、音声認識誤りにより、検索に必要な情報が欠損する可能性もある。このような場合には、検索の手がかりとなる情報が少なくなるため、結果として非常に多くの候補がマッチング結果として得られることになる。そのため、ユーザに聞き返しを行い、不足している情報や音声認識誤りにより失われた情報を補完することで、候補を絞り込む必要がある。

本研究では、ある程度ドメインが限定された文書検索タスクを対象として、検索結果の絞り込みを行うために対話的に検索要求を明確化する手法を提案する。まず、知識ベースから抽出されるボトムアップの知識と、人間の知見などを利用したトップダウンの知識を併用することにより、可能な質問集合を用意する。その際に、検索結果に対する情報ゲインが大きくなるように動的に質問を選択することで、効率のよい対話を実現する。さらに、生成した質問に対するユーザの回答を用いて、検索文を更新・整形することも考える。

表 1: ソフトウェアサポート用知識ベース

知識ベースの種類	件数	文字数
用語集	4,707	約 70 万
ヘルプ集	11,306	約 600 万
サポート技術情報	23,323	約 2,200 万
合計	39,336	約 3,000 万

## 2 音声による大規模文書検索システム

本研究で検索対象とする文書は、マイクロソフト社のソフトウェアサポート用知識ベースであり、その概要を表 1 に示す。これらは自然言語によって記述されている。

これまで我々は、音声による検索要求によりこの知識ベースを検索するシステムとして「音声版ダイアログナビ」を開発してきた [3]。音声版ダイアログナビでは、ユーザ発話の音声認識結果に対して文節ごとに知識ベースとの整合性の尺度となる検索整合度、音声認識結果の N-best 候補により計算される検索重要度により、検索の前後で確認を行い音声認識誤りや、音声言語の冗長性に対処してきた。しかし、従来のシステムは基本的に一問一答型のシステムであり、漠然とした検索要求や、音声認識誤りにより一部の文節が削除された場合には、多数の候補がマッチングし、そのまま提示されていた。今回、この音声版ダイアログナビを拡張し、検索要求を明確化するための対話戦略を実装する。

## 3 漠然とした検索要求を明確化するための対話戦略

本研究では可能な質問集合を用意しておいて、検索結果に基づいて動的に選択する枠組みを考える。前提条件を満たす質問の中から最も効果的なものを選択す

るために、その質問により得られる情報ゲインを定義する。これは、該当する文書の候補数をどれくらい絞り込める（分類できる）かという意味でエントロピーに基づいた量である。ただし、ある質問に対して複数のカテゴリに属するテキストが存在したり、逆にどのカテゴリにも属さないものが存在するため、単純なエントロピーを適用することはできない。そこで、以下の式により質問  $S$  による情報ゲイン  $H(S)$  を定義する。

$$H(S) = - \sum_{i=0}^n P(i) \cdot \log P(i)$$

$$P(i) = \frac{|C_i|}{\sum_{i=0}^n |C_i|}$$

$$|C_i| = \sum_{D_k \in i} CM(D_k)$$

ここで、 $D_k$  は最初の検索要求でマッチングした  $k$  番目の候補であり、 $CM(D)$  は文書  $D$  のマッチングのゆう度である。 $C_i$  は質問  $S$  で分類されるカテゴリ  $i$  (計  $n$  個) に該当する文書の数に相当するが、マッチングのゆう度で重み付けされている。また、どのカテゴリにも該当しないテキストはカテゴリ 0 に属するものとする。

この質問戦略を組み込んだ処理の流れは以下の通りである。

1. 音声認識結果に対して検索整合度を計算し、文節に対して検索重みを付与する。これにより、認識誤りや検索に不要な箇所を除去する。
2. ダイアログナビ [4] を用いて知識ベースを検索する。
3. 全ての質問候補に対して、情報ゲインを計算する。
4. 情報ゲインが大きく、候補の絞り込みが期待される質問が存在する場合には、その中から最も情報ゲインが大きい質問を生成する。それ以外の場合には、ユーザに最終結果を提示する。
5. 質問に対するユーザの回答を用いて、検索文を更新する。
6. 2に戻る。

これら全体の流れを図 1 に表す。また、次章以降でその詳細について述べる。

## 4 ボトムアップの知識とトップダウンの知識を利用した質問生成

本研究では、マッチングした文書テキストの文から得られるボトムアップの知識と人間の知見や知識ベ

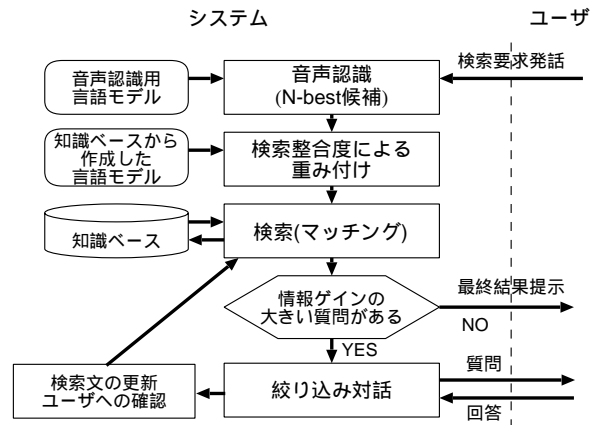


図 1: 絞り込み戦略の概要

スから抽出できるトップダウンの知識の両方を利用して、以下の3つの手法により質問候補を用意する。このうち、ボトムアップな知識を利用した質問として「係り受け情報を利用した質問」、トップダウンの知識を利用した質問として「人間の知見を利用した質問」、知識ベースのメタデータを用いる質問」を考える。

### 4.1 係り受け情報を利用した質問 (手法 1)

検索要求に含まれる単語を係り受け対象とする語句が明示的に示されていない場合に、それを明確化するための質問を行う。例えば、「削除する」という単語を含む候補には「アプリケーションを削除する」や、「アドレス帳を削除する」のように数多くの候補が存在する。そこで、これらの単語の係り元を同定することで、検索要求を明確化することができる。ただし、その際に全ての単語について確認すればよいわけではない。例えば「コンピューターシャットダウン」という係り受けペアのように係り元の語句がほぼ一意に同定できる語句も存在し、そのような語句に対して質問するのは冗長である。そのため、どのような単語を対象として質問を生成するかは、コーパスから係り受けペアに関するエントロピーを計算することにより決定する。具体的な手順は以下の通りである。

1. 知識ベース、及び「マイクロソフト話し言葉検索」に寄せられた検索文から全ての係り受けペアを抽出する。
2. 単語  $m$  に係る語句が単語  $i$  である頻度割合  $P(i)$  を求め、それによりエントロピー  $H(m)$  を計算する。

$$H(m) = - \sum_i P(i) * \log P(i)$$

このエントロピー  $H(m)$  が大きい上位 40 単語を抽出して確認することとした。

表 2: システムが生成する質問の例

質問の内容	手法が適用可能な知識ベースの割合	情報ゲイン
係り受け情報を利用した質問 (手法 1)		
何を削除したか	2.15 (%)	7.44
何をインストールしたか	3.17 (%)	6.61
人間の知見を利用した質問 (手法 2)		
症状はいつ発生したか	15.40 (%)	8.08
発生したエラーの内容	2.63 (%)	8.61
知識ベースのメタデータを利用した質問 (手法 3)		
Windows のバージョン	30.03 (%)	2.63
Application の種類	30.28 (%)	2.31

## 4.2 人間の知見を利用した質問 (手法 2)

本研究で対象とするソフトウェアサポートは、コールセンターのオペレータによっても行われている。そこにおいて蓄積されている人間の知見に基づいて質問候補を用意する。今回、この手法による質問を 3 種類用意した。例えば、「その症状はいつ発生しましたか?」という質問では、「~した際」「~した後」という文節の直前に現れる名詞を同定するものとする。

## 4.3 知識ベースのメタデータを用いた質問 (手法 3)

検索対象の知識ベースに付与されているメタデータを利用して質問を生成する。一般的に、新聞記事や図書のような大規模な資料や文書を扱う際には、管理・運用を効率的に行うために、各文書にカテゴリ情報などのメタデータが付与されることが多い。本研究の検索対象のソフトウェアサポート用知識ベースにおいても、その文書がどのプロダクトについて記述したものであるかを表すメタデータが付与されているものがある。このような情報を利用して、ユーザの検索要求がどれに該当するかをたずねる質問を生成する。本研究で使用する質問の数は 14 である。

以上の 3 つの手法によって生成する質問の例を表 2 に示す。また、実際の検索結果に対して情報ゲインを計算し、システムが質問を生成する対話例を図 2 に示す。

## 5 システムの質問に対するユーザの回答を利用した検索文の更新

このように生成した質問に対するユーザの回答を利用して単純に候補を絞り込むこともできるが、最初のユーザ検索発話を更新・整形し、再度マッチング

S1:	ご質問をどうぞ。
U1:	文字化けしました。
(検索結果):	1. [閉じる] ボタンや [最大化] ボタンが文字化けした場合の対処方法 2. [OL 2000] Outlook Today 画面の文字化け 3. テキストの挿入処理で、日本語入力を行なうと文字化けする 4. [ XL2002 ]ピボットテーブルを作成するとドロップダウンリストのアイテム名が文字化けする 5. VB アプリが文字化けする ...
(情報ゲイン計算):	質問候補 1: 「何が文字化けしましたか?」 (情報ゲイン 5.27) 質問候補 2: 「その症状はいつ発生しますか?」 (情報ゲイン 2.47) 質問候補 3: 「Windows のバージョンを教えてください」 (情報ゲイン 1.43) ...
S2:	(最も情報ゲインが大きい質問を選択) 何が文字化けしますか?
U2:	ウインドウのボタンです。
S3:	(検索文の更新) 「ウインドウのボタンが文字化けする」 で検索します。

図 2: 対話の例

を行う方が、検索文中の語句の係り受け関係などの情報を利用してより正確なマッチングができると期待される。また、最終的な検索要求をユーザに確認する際に、システムの質問への回答を羅列して確認するよりも更新した検索文を提示する方がわかりやすいと考えられる。そこで、質問に対するユーザの回答を用いて最初の検索発話を更新する。本研究では、質問手法ごとにあらかじめ検索文の更新規則を用意する。それぞれの質問手法ごとの更新規則は以下の通りである。

1. 係り受け情報を利用した質問  
ユーザの回答を係り先となる単語の直前に追加する。例えば「何を削除するのですか?」という質問に対する回答は、検索文中の「削除」という単語の直前に挿入する。
2. 人間の知見を利用した質問  
生成する質問ごとに、ユーザの回答を挿入する箇所をあらかじめ指定しておく。例えば「エラーはいつ出ますか?」という質問に対する回答は検索文の「エラー」という単語の直前に挿入する。
3. 知識ベースのメタデータを利用した質問  
ユーザの検索文の文頭に「(タグ名)で」の文節を付加する。また、プロダクトのバージョンを確認する質問の場合は、検索文の「(プロダクト名)」の部分に「(プロダクト名)(バージョン情報)」に置き換える。

## 6 被験者実験による対話戦略の評価

提案する対話戦略の有効性を確かめるために、14名による被験者実験を行った。被験者には、想定場面に基づく14課題について検索を行ってもらった。ただし、回答としてふさわしい検索結果が得られない場合には、被験者の判断で各課題につき1度のみ言い直しを許した。その結果、合計196課題、238発話のデータを得た。

まず、実際の対話のログデータによりシステムの評価を行った。提案手法による検索成功率と、正解候補を提示した順位について、人手により作成したユーザ発話の正確な書き起こし、及び音声認識結果をそのまま利用して検索を行った場合との比較を行った。これに加えてシミュレーションにより、提案する検索要求の明確化手法のみの効果も検証した。対話のログデータによる評価は、インターフェースを含むシステム全体の種々の要因が影響している。具体的には、検索整合度によるマッチングによる効果も含まれている反面、システムの生成する質問に対して、ユーザが的確に回答しているとは限らない。そのため、システムからの質問に対して的確な回答が得られるという前提のシミュレーションにより評価を行った。これらの結果を表3に示す。

音声認識結果でそのまま検索を行った場合と比べて、実際のシステムを使った実験では7.7%、シミュレーションによる評価では12.6%検索成功率が向上している。シミュレーションによる検索成功率の方が約5%高い原因を分析したところ、質問への回答発話の音声認識誤りや、質問意図に対するユーザの誤解が主な原因であることがわかった。

さらに、質問手法ごとの効果を調べるために、それぞれの手法のみを用いた場合の検索成功率と、その際に生成された質問の回数を求めた。この結果を表4に示す。各手法によるそれぞれの検索成功率には大きな差はなく、また向上の幅もそれほど大きくないが、3つの手法を併用することによる相乗効果により、高い検索成功率が得られている。ただし、係り受け情報を利用した質問(手法1)が質問回数が少ない割に改善幅は他と同程度になっており、最も効率がよい手法といえる。

## 7 むすび

本研究では、ソフトウェアサポート用知識ベースを対象とした話し言葉音声による文書検索タスクにおいて、ユーザの漠然とした検索要求に対して対話的に意

表 3: ユーザの実際の対話によるシステム全体の評価

評価対象	検索成功率	正解候補の平均順位
書き起こし入力(参考)	76.1%	7.20
音声認識結果入力(ベースライン)	70.7%	7.45
提案手法(対話ログの分析)	78.4%	4.40
提案手法(事後評価)	83.3%	3.85

表 4: 質問手法ごとの効果の分析

	検索成功率	確認回数
ベースライン(参考)	70.7%	—
係り受け情報(手法1)	74.5%	0.38
人間の知見(手法2)	74.5%	0.97
メタデータ(手法3)	75.7%	0.89
全手法(手法1~3)	83.3%	2.24

図を明確化するための対話戦略を提案した。そのために、係り受け関係を利用した質問、人間の知見を利用した質問、知識ベースのメタデータを利用した質問を用意し、情報ゲインの基準により選択する枠組みを採用した。さらに、その回答により検索文を更新することも行った。14名の被験者による評価実験を行ったところ、音声認識結果そのままでは検索が失敗した発話の42.9%に対して、対話により回答となる知識ベースを提示することができた。

## 謝辞

本研究に対し、多大な協力を頂いた東京大学の黒橋禎夫助教授、清田陽司助手、及びマイクロソフト株式会社の木戸冬子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, 木戸冬子. 音声対話による大規模知識ベース検索システム -音声版ダイアログナビ-. 情報処理学会研究報告, 2004-SLP-52-4, 2004.
- [4] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答 -ダイアログナビ-. 自然言語処理, Vol. 10, No. 4, pp. 145–175, 2003.