

論文 / 著書情報  
Article / Book Information

論題(和文)	音声インタラクティブODQAの構築とその評価
Title(English)	
著者(和文)	堀智織, 堀貴明, 磯崎秀樹, 前田英作, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2003年春季講演論文集, Vol. , No. 2-4-7, pp. 71-72
Citation(English)	, Vol. , No. 2-4-7, pp. 71-72
発行日 / Pub. date	2003, 3

堀 智織 堀 貴明 磯崎 秀樹 前田 英作 (NTT コミュニケーション科学基礎研究所)  
古井 貞熙 (東工大)

## 1. はじめに

本研究では、音声インタラクションに基づく質問応答プロセスを組み込んだ ODQA (Open Domain Question Answering) システムの枠組みを提案する。近年、新聞記事などの大規模コーパスを対象とした ODQA システムの研究が盛んに行われている [1]。提案する枠組は従来のテキスト入力による一問一答形式とは異なり、ユーザーの音声による断片的かつ曖昧な質問からシステムが必要に応じて確認や問い返しを行い、正しい解答を導く機能を持つ。本稿では、我々が構築している音声インタラクティブ ODQA システム SPIQA (SPoken Interactive QA) の概要について述べ、ユーザに問い返しを行う Deriving Disambiguous Queries (DDQ) モジュールにより自動生成させた情報要求文を評価した結果を報告する。さらに、この情報要求文に基づきユーザの質問文に新たな情報を付加したところ、QA の正答率の向上を確認した。

## 2. SPIQA の構成

音声インタラクティブ ODQA システム SPIQA の基本構想を図 1 に示す。ユーザの発声した質問音声を音声認識装置で認識し、認識誤りを排除しつつ言語的に尤もらしい単語列を抽出し ODQA に入力する。ODQA から得られた解答に対し、確信度が高い場合には解答を出力、確信度が低い場合にはユーザに情報補完をさせる情報要求文を自動生成する。システムによる情報の要求とユーザによる情報の補完に基づくインタラクションにより、質問に対する解答を抽出する。SPIQA システムは音声認識システム、スクリーニングフィルタ、情報要求を行わない ODQA システム (SAIQA)、解答候補評価部、曖昧性解消のための自動情報要求文生成装置により構成される。

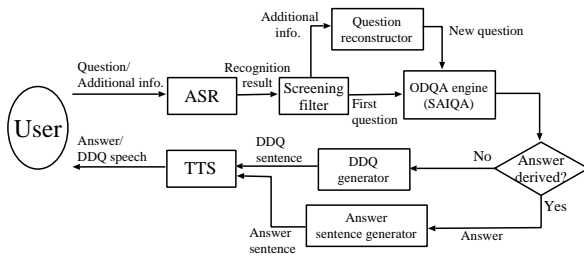


図 1. 音声インタラクティブ QA : SPIQA の基本構成

### 【音声認識装置】

音声認識装置として、有限状態トランスデューサを用いる [2]。音響モデルおよび言語モデルを統合したコンパクトな状態ネットワークに基づき効率的な音声認識を行なう。

### 【スクリーニングフィルタ】

音声認識結果には、人間による発話の誤りと認識装置による誤りという二種類の誤りが含まれる。従って、認識結果をそのまま ODQA システムに入力した場合、解答抽出の性能は著しく劣化する。この問

題を解決するため、質問文の音声認識結果から音響的および言語的に尤もらしい単語列を抽出するスクリーニング処理を行う。これは、音声認識結果から信頼度の高い重要な単語を言語的に尤もらしくなるように単語抽出する音声自動要約手法 [3] の枠組を適用する。

### 【情報要求の無い ODQA エンジン: SAIQA】

システムからの情報要求の無い ODQA システムの一般的なシステムとして SAIQA [4] を用いる。SAIQA は自然文テキストによる質問文を入力とし、その解答を新聞記事から検索するシステムである。具体的には、入力された質問文に対し、ユーザが欲している情報が何であるかを質問タイプとして同定し、さらにキーワードを抽出する質問解析を行う。次に、抽出されたキーワードと質問タイプに基づき検索対象となるコーパスから関連テキストを検索する。検索されたテキスト中において、質問タイプと対応づけられる名詞句を同定し解答候補を抽出する。さらに、解答としての信頼度に基づき、解答候補の中から解答を決定し出力する。

### 【情報要求文生成装置: DDQ モジュール】

質問文中に解答抽出に必要な情報が十分含まれていない、または音声認識により重要な情報が欠落している場合、ODQA エンジンは解答を正確に抽出できないことがある。このように解答抽出に十分な情報が得られなかった場合、質問文には「曖昧性」があると定義する。この曖昧性を解消するため、解答抽出に必要な情報をユーザに要求する情報要求文を DDQ モジュール [5] により自動生成する。スクリーニング処理された認識結果と検索対象における単語の出現傾向に基づき、質問文の曖昧箇所を推定する。推定された曖昧箇所を疑問文のテンプレートと組み合わせ、疑問文として尤もらしい文を情報要求文として選択する。以下に質問文の曖昧性を計る尺度について説明する。

### 構文的曖昧度

質問文において他の文節から係られていない文節は、意味が限定されておらず曖昧性が高いと考えられる。そこで、当該文節が他の文節から係られている度合いに基づく曖昧度を構文的曖昧度とする。質問文内の  $n$  番目の文節の構文的曖昧度スコア  $A_D(P_n)$  を次式のように定義する。

$$A_D(P_n) = \log \left\{ 1 - \sum_{i=1; i \neq n}^N D(P_i, P_n) \right\}$$

但し、 $D(P_i, P_n)$  は  $P_n$  番目の文節が  $P_i$  番目の文節により係られている確率と定義する。本研究では、SD-CFG (Stochastic Dependency Context Free Grammar) [3] に基づき、 $D(P_i, P_n)$  を計算する。但し、 $k \geq l$  の場合、 $D(P_k, P_l)$  は 0 である。

\* Construction of Spoken Interactive ODQA system and Its Evaluation

## 単語汎用度

検索対象において出現頻度の高い単語は、解答抽出において候補の絞りこみに貢献しないことから、曖昧性が高いと言える。そこで、実質語に対する曖昧性を次式で示す汎用度スコアにより次式のように定義する。

$$A_G(P_n) = \sum_{w \in P_n: w=\text{cont}} \log P(w),$$

但し、 $P(w)$  は検索対象となるコーパスにおける  $w$  の unigram 確率とする。“ $w = \text{cont}$ ” は、 $w$  が名詞、動詞、または形容詞といった実質語である場合を示す。 $S_{mn}$  は、ユーザの質問文の  $n$  番目の文節と  $m$  番目の質問文テンプレートから自動生成した情報要求文である。DDQ モジュールは次式で示す DQ スコアを最大とする情報要求文を生成する。

$$H(S_{mn}) = \lambda_L L(S_{mn}) + \lambda_D A_D(P_n) + \lambda_G A_G(P_n),$$

ただし、 $L(\cdot)$  は言語スコアを示す。本研究では trigram 確率の対数値を適用する。 $\lambda_L, \lambda_D, \lambda_G$  は、スコアのバランスを取るための重み係数である。

## 3. 評価実験

### 3.1 実験条件

7人の男性被験者が読み上げた69文の質問音声を音声認識装置により認識し、DDQ モジュールにより情報要求文を自動生成し評価した。読み上げた質問文は、文法的に正しく書き言葉で書かれた文である。平均単語正解精度は76%であった。

### 音声認識部

話し言葉コーパスから構築された、各状態16混合ガウス分布の不定話者音素文脈依存 HMM(3000 状態)の音響モデル [6] を適用した。毎日新聞1994年～2001年までの8年分をJTAGにより形態素解析し、語彙サイズ2万で作成した bigram, trigram を言語モデルとして適用した。

### DDQ モジュール

単語の汎用度スコア  $I$  は、毎日新聞8年分を適用した。質問文作成に用いた疑問詞は、「誰の」「何の」等、これまでSAIQAに入力された疑問詞82個を適用した。質問文としての適正を示す言語モデル  $L$  には、毎日新聞3年分と質問文1000を併せて学習した trigram を適用した。係り受け確率  $D$  については、京大コーパスに基づき学習したSDCFG [3] を適用した。

### 3.2 自動応答対話文の評価方法

質問文の認識結果と正解書き起こし文に基づき、5人の被験者が作成した情報要求文を正解とした。この正解情報要求文とDDQモジュールが自動生成した情報要求文の比較評価を行った。自動生成された情報要求文が正解と同じ内容の質問の時、適正であると評価した。但し、ユーザの入力した情報の全てがQAに不可欠なものではない。そこで、DDQモジュールがユーザに補完を要求した情報が、QAの解答抽出において必要とされる情報であるかを調べた。具体的には、被験者が質問文のスクリーニング結果に情報要求文により要求された新しい情報を付加し再構成した。ただし、適正な情報要求文 (APP) のみを使用した。再構成された質問文を用いて、再度QAによる解答抽出を行い性能の改善を調べた。QAの性能は、MRR (Mean Reciprocal Rank) [7] を用いて評価した。MRRは正解の順位の逆数の平均で

表1. 自動応答対話文の例

読み上げ質問	絶滅の危機に瀕するマウンテンゴリラは、どこにいますか？
音声認識結果	絶滅の危機に瀕する満点これらは、どこにいますか？
スクリーニング結果	絶滅の危機に瀕するこれらは、どこにいますか？
自動応答対話文	何の絶滅ですか？

表2. DDQにより生成された情報要求文の評価結果

SPK	Word acc.	MRR			w/o errors	APP	In-APP
		REC	SCRN	DQ			
A	70%	0.19	0.20	0.23	4	32	33
B	76%	0.31	0.28	0.31	8	36	25
C	79%	0.25	0.26	0.30	10	34	25
D	73%	0.28	0.27	0.30	4	35	30
E	78%	0.26	0.24	0.27	7	31	31
F	80%	0.30	0.29	0.33	8	34	27
G	74%	0.19	0.19	0.22	3	35	31
AVG	76%	0.25	0.25	0.28	9%	49%	42%

MRR以外で%でない実数値は文数を示す。Word acc.: 単語正解精度, SPK:話者, REC: 認識結果, SCRN: スクリーニング結果, DQ: DQに基づく補完, AVG: 平均値, w/o errors: 文正解, APP: 適正なDQ, InAPP: 不適切なDQ。

ある。本実験では5位以内の正解に対し、MRRを計算した。

### 3.3 実験結果

自動情報要求文の例と評価結果を表1と表2に示す。実験結果より、認識誤りの含まれる質問に対し、約5割の質問で5-best中に曖昧性を解消する応答対話文が生成できたことが分かる。69文の平均MRRは、正解書き起こし文、音声認識結果 (REC)、スクリーニング結果 (SCRN)、情報要求文に基づき再構成された質問文 (DQ) を用いた場合、各々0.43, 0.25, 0.25, 0.28であった。これにより、情報要求文に基づき補完された情報によるQAの性能改善が示された。

## 4. まとめ

音声インタラクションに基づく質問応答プロセスを組み込んだODQA (Open Domain Question Answering) システムの枠組みを提案した。誤認識を含んだ質問文をスクリーニングし、情報要求文を自動生成した結果、約5割の質問文で5-best中に適正な情報要求文を自動生成できた。さらに、自動情報要求文で要求された情報を補完したところ、QAの性能が改善された。今後、コーパス中の単語間の係り受け関係や意味カテゴリを考慮し、曖昧性解消の性能を改善したい。さらに、情報要求を繰り返し効率的な情報検索を行うための対話戦略の枠組みを検討したい。

### 参考文献

- [1] <http://trec.nist.gov>
- [2] D. Willett et. al., *Proc. of Eurospeech 2001*, Vol. 2, pp. 847-850, 2001.
- [3] C. Hori et.al., "A New Approach to Automatic Speech Summarization," To appear in the *IEEE Transactions on Multimedia*, 2003.
- [4] Y. Sasaki et. al., *Proc. of NTCIR Workshop Meeting*, pp. 63-70, 2000.
- [5] 堀他, 音講論, Vol.1, 2-9-11, pp.83-84(2002).
- [6] 篠崎他, 音講論, Vol.1, 1-3-14, pp.31-32(2001).
- [7] <http://trec.nist.gov/data/qa.html>