

文献検索音声対話システムの機能拡張とその評価

桐山 伸也 広瀬 啓吉
東京大学 工学部

学術文献検索をタスクとした音声対話システムについてシステムの機能を拡張し、構築した新システムを評価した。以前に構築したシステムの評価結果を基に、音声認識部の性能向上を図るとともに、(1)GUIの有効利用、(2)より高度な意味処理の導入の2点を柱とした機能拡張を施すことによって、よりユーザフレンドリなシステムへと改良した。以前のシステムと新システムの両方を実際に被験者に使用してもらい、両者を比較することで新システムを評価した。その結果、新システムの拡張機能の有効性が確かめられた。

Functional Extension of Spoken Dialogue System for Academic Document Retrieval and its Evaluation

Shinya KIRIYAMA Keikichi HIROSE
Faculty of Engineering, University of Tokyo

We evolved a spoken dialogue system for academic document retrieval by extending its functions, and evaluated the system. In order to reform the system more friendly to users, we introduced a new speech recognizer to improve recognition performance, while enlarged functions by considering two points: 1) to use GUI effectively, 2) to import more intelligent semantic processing. Evaluation was conducted by asking six users to compare the new and previous systems. The results indicate the extended functions were valid, though the system still has a room for improvement to deal with misunderstanding or detailed semantic contents.

1 はじめに

我々は従来から、学術文献検索を音声によって行なうシステムを開発中である。ユーザとの円滑な対話を進めていくにあたり、適度に確認や状況説明を交えながら、システムが対話の各時点で適切な応答をすることに主眼を置いて研究を進めており、この観点から構築したシステムの評価を行なっている [1]。

今回この評価実験の結果を基に、新たな音声認識器の導入とシステムの機能拡張を図り、新システムを構築した。

さらに、構築したシステムと以前のシステムの双方を実際に稼働させて被験者に試用してもらい、従来システムとの違いを踏まえた上で、新システムの評価を行なった。

以下、まずシステムの構成を述べ、次いでシ

ステムの機能について従来機能と拡張機能を説明し、最後に評価実験について述べる。

2 システムの構成

本システムは、音声認識部・対話管理部・検索部・画面表示部・音声合成部の5つのモジュールからなる。図1に本システムの構成を示す。

以下、各構成要素について述べる。

2.1 音声認識部

従来、HTK [2] によって音声認識器を作成して用いていた。対話の各時点でのシステムの状態によって文法を切り替えることにより認識率の向上を図っていたが、処理速度が遅かったためにあまり良い評価を得られなかった。

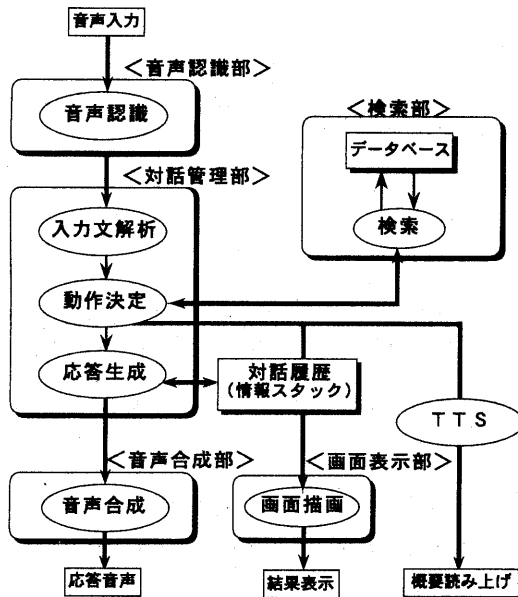


図 1: システムの構成

今回は速度向上と機能拡張に伴う語彙・文法の拡張に対処すべく、情報処理振興事業協会 (IPA) の独創的情報技術育成事業の研究成果物である「日本語ディクテーション基本ソフトウェア 98 年度版」における Julius [3] の文法ベースバージョンにあたる Julian [4] を音声認識エンジンとして使用した。

音韻モデルについては、上述のソフトウェアに含まれる、状態数 1,000・混合数 4 の triphone モデルを利用した。

言語モデルについては、従来 HTK で用いていた単語辞書と構文規則を基に、今回の機能拡張に伴って増加した分をそれぞれ追加し、これを Julian 附属のツールを用いて、Julian の指定する形式に変換して用いた。

この結果、語彙は検索語が 375 語、機能語が 160 語の合計 535 語となった。因みに従来の HTK での語彙は検索語同数、機能語 125 語の計 500 語であった。

2.2 対話管理部

認識結果を解析してシステムの動作を決定し、応答を生成するモジュールである。入力

「検索して下さい」や「表示して下さい」といったコマンドならその旨を実行すべく、検索部・画面表示部に指令を送る。

本システムでは、状態遷移表に基づく対話管理方式を採用している。この方式に則った場合、条件入力状態から検索状態に移行するタイミングや、検索直後の検索結果の出力方法などは一意に定まるものではなく、幾通りか方法が考えられる。

以前構築したシステムにおいては、これらを制御する方式を複数用意し、実験を通してこれらの評価することを行なった。その結果、システム主導とユーザ主導の中間的な動作モードが好まれるとの結果を得たため、今回のシステムではこのモードをデフォルトとして採用することとした。

2.3 検索部

入力された検索語と各検索語間の検索方式 (AND 検索または OR 検索) の情報から検索文字列を生成し、文献データベースにアクセスして検索を行なう。

文献データとしては、学術情報センター電子図書館サービス [5] のものを使用している。日本音響学会誌・電子情報通信学会論文誌等、約 1,200 本の文献データから検索を行なう。

2.4 画面表示部

従来はコンソール画面にテキスト情報のみを表示するという単純な仕組みの表示部を便宜上設けていたが、今回の機能拡張の 1 つの柱として、画面表示部を独立に用意して用いることとし、Tcl/Tk によって実装を行なった。

本表示部は、対話管理部からの指令を受けて表示内容を更新する。表示できる内容は、認識結果・検索文字列・該当件数・タイトル一覧・文献詳細内容・選択単語リストの 6 カテゴリに分類され、状況に応じて取捨選択的にこれらの一部ないし全部が表示される。

2.5 音声合成部

対話管理部が生成した単音・韻律記号列を入力とし、ターミナルアナログ方式の音声合成器を起動して応答音声を出力する。

なお今回、韻律情報辞書のアクセント型の誤りなどを修正し、正しい指令に則って音声を合成するようにした。

3 システムの機能

本システムは、1) 検索語によって文献検索を行なう機能、2) 検索結果を GUI により表示し文献の概要を読み上げる機能、3) 文献情報に関する質問に答える機能、という大きく分けて3つの機能を持つ。

今回システム全体に対して機能拡張を図ったわけであるが、その主眼点は次の2点である。

- 新たに実装した GUI への移行に伴う操作性の向上
- タスク達成支援に向けたより多様な意味処理の導入

以下上述の3機能について、従来からの機能の説明に加え今回新たに実装した拡張機能について具体的に述べるとともに、システムの動作例を示す。

3.1 検索機能

検索語をタイトル・キーワード・アブストラクト・著者名・発行年・掲載誌名の6カテゴリとこれらに跨るカテゴリの計7カテゴリに分類した上で、各カテゴリの内部及びカテゴリ間で、AND 検索と OR 検索を組み合わせる検索を行なう機能を有していた。

今回新 GUI の採用により、検索文字列の操作を容易に行なえるようにした。これにより、「2番目の検索語をタイトルで検索して下さい」、「1番目の AND を OR に変更して下さい」といった操作が可能となった。さらに検索文字列の変化に対応して、変更が加わった部分だけを

強調色で表示することにより、直前の検索文字列との違いを明確に示せるようにした。

また、現在の検索語に関連する検索語のリストを提示できるようにした。これは該当文献数が多い場合に、そこから見通し良く文献の絞り込みを行なっていくにあたり有効な機能と考えられる。

3.2 結果出力機能

従来、文献のタイトル一覧を表示する簡略表示機能、著者名・掲載誌・概要をも表示する詳細表示機能に加え、TTS システムによって文献のアブストラクトを読み上げる機能が実装されていた。

GUI の移行によって表示画面が見易いものとなった。表示する情報の変化に伴い、その時点で新たに変更ないし追加された項目を強調色で表示する機構を実装している。

また以前に表示した内容を再表示する操作ができるようになった。これは「1つ前のものを表示して下さい」といった入力によって対処できる。

3.3 質問応答機能

検索結果に対して、文献のタイトル・著者名・発行年・掲載誌名と、アブストラクトの有無の5項目についての質問に音声で応答することが可能であった。

上述のように従来は各々の文献の詳細に関わる項目しか質問できなかったが、今回該当文献全体に対する質問を受け付けられるようにした。

すなわち、「この中で一番新しい文献はどれですか」、「一番多くの文献が掲載されている雑誌名を教えてください」といった内容の質問が可能になった。

3.4 システム動作例

以下に、システム動作例として図2に対話の例を、図3に表示画面例を示す。

- S-1 どのような論文をお探しですか？
 U-1 音声認識
 S-2 現在の条件で検索しますか？
 U-2 はい
 S-3 では検索します。
 検索した結果 24 件が該当します。
 もう少し条件を絞りますか？
 U-3 はい
 S-4 条件を入力して下さい。
 U-4 HMM
 S-5 では現在の条件で検索しますか？
 U-5 はい
 S-6 検索した結果 以下の 8 件が該当します。
 どれかについてアブストラクトを
 表示しますか？
 U-6 はい
 S-7 何番のアブストラクトを表示しますか？
 U-7 1 番
 S-8 1 番のアブストラクトを表示します。
 U-8 2 番を詳しく表示して下さい
 S-9 2 番のアブストラクトを表示します。
 U-9 1 つ戻って下さい
 S-10 以下に 1 つ前を表示します。
 U-10 一番新しいものは何ですか
 S-11 7 番・8 番が 1997 年で一番新しいです。
 U-11 7 番の著者名は何ですか
 S-12 南泰浩です。
 U-12 8 番は誰ですか
 S-13 著者は奥乃博・川端豪・中谷智広です。
 U-13 雑誌名は何ですか
 S-14 8 番は情報処理学会論文誌です。
 U-14 もう結構です
 S-15 終了しますか？
 U-15 はい
 S-16 御利用ありがとうございました。

図 2: 対話例

4 評価実験

構築した新システムと以前開発したシステムを実際に稼働させて評価実験を行なった。

4.1 評価の方法及び項目

評価は、旧システム・新システムの順に 2 つのシステムを被験者に試用してもらうことにより行なった。

まず音声認識器の性能について、単語認識率と文理解率を算出した。そして、被験者に対しては、新旧両システムの違いを踏まえた上で、項目別のアンケートに回答してもらうと同時に

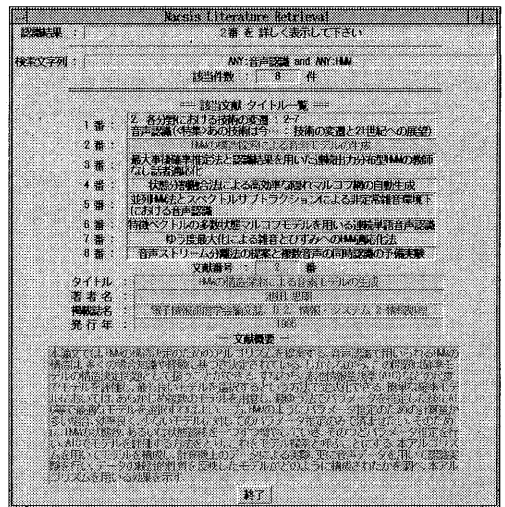


図 3: 詳細表示の例

にコメントを記入してもらい、新システムへの旧システムに対する相対的な評価を行なった。

アンケートについては、各項目について中間を 3 とした 5 段階評価を記入する形を取った。アンケート及びコメント記入方式で評価した項目は、以下の 5 つである。

- 音声入力
認識時間は短縮されたか、など 2 項目。
- 画面表示
画面の色の変化はどうか、など 6 項目。
- 諸機能
前画面表示機能は有効か、など 4 項目。
- 応答音声
応答音声は聞き取り易いか、など 3 項目。
- システム全体
使い勝手は良かったか、など 4 項目。

4.2 実験結果

まず、音声認識器についての結果を表 1・2 に示す。ただし、単語認識率 WCR 、文理解率 SUR は以下の式で定義する。

表 1: 単語認識率

Testee	W_{all}	W_{sub}	W_{ins}	W_{del}	WCR
A	69	13	2	7	68.1
B	35	7	2	0	74.3
C	24	6	0	0	75.0
D	45	7	3	2	73.3
E	49	13	0	2	69.4
F	77	14	0	4	76.6
T_{new}	299	60	7	15	72.6
T_{old}	221	43	1	3	78.7

表 2: 文理解率

Testee	S_{all}	S_{und}	SUR
A	37	31	83.8
B	22	17	77.3
C	17	11	64.7
D	28	22	78.6
E	35	29	82.9
F	37	32	86.5
T_{new}	176	142	80.7
T_{old}	140	108	77.1

$$WCR(\%) = (W_{all} - W_{sub} - W_{ins} - W_{del}) / W_{all}$$

$$SUR(\%) = S_{und} / S_{all}$$

ここで $W_{all} \cdot W_{sub} \cdot W_{ins} \cdot W_{del}$ は、対話中の単語全て・置換誤り・挿入誤り・脱落誤りの数を、 S_{all} と S_{und} はユーザの発話の総数・システムがユーザの発話内容を正しく理解した発話の数をそれぞれ表す。また A ~ F は新システムにおける被験者個別の数値、 T_{new} 、 T_{old} はそれぞれ新システム・旧システムにおける被験者全体の数値である。

文理解率は今回の Julian の方が、以前の HTK のものより良かったものの、単語認識率は逆に低下している。これは今回の認識器が、検索語の入力に際して「音声認識と HMM」のように複数の検索語を一度に入力できる文法に変更したことに起因すると考えられる。検索語を検索条件としてシステムに入力する際の認識誤りは、従来の HTK のものより少ないにも関わらず、検索条件を変更する際の、「(検索語)を消して下さい」といった発声に対して、「(検索語)と(検索語)」という誤認識が多発していた。(従来の文法では、この形の文を受理しないため、この種の誤りは先頭の検索語が誤っているものしかあり得なかった。)こ

表 3: アンケート集計結果

評価項目	評価
音声入力	
旧システムと比べて認識時間は	2.5
旧システムと比べて認識率は	0.0
画面表示	
旧システムと比べて見易くなったか	6.7
新システムの画面の切替えはどうか	5.8
新システムの認識結果の表示は	10.0
新システムの関連語リストの表示は	8.3
新システムの色が変化するの	5.8
新システムの検索文字列の部分強調は	5.8
語機能	
新システムの前面表示機能は	5.8
新システムの一番多い文献の機能は	-0.8
新システムが一番新しい文献の機能は	6.7
新システムの検索文字列の操作機能は	5.0
応答音声	
旧システムと比べて合成音声は自然か	3.3
旧システムと比べて聞き取り易いか	4.2
旧システムと比べて強調がわかるか	1.7
システム全体	
旧システムと比べて処理速度は	3.3
旧システムと比べて機能充実したか	5.8
旧システムと比べて使い勝手は良いか	5.0
新システムをまた使いたい	5.8

の誤りが1つ生じる度に誤った検索語が検索文字列に挿入されることになり、その単語を消去しようとしてさらにまた新たな誤検索語を生じさせてしまうという悪循環に陥ってしまった例もあった。また今回は、検索条件の変更に際してより柔軟な操作を許したため、この種の発話数が増加したことも認識率低下に迫車をかけたものと考えられる。これに対しては、検索文字列に入力検索語を追加する際に一段確認の対話を設ける等の対処が必要になると思われる。

次に、アンケートの結果を表3に示す。数値は5段階評価の高い方から順に+10・+5・0・-5・-10ポイントを与えるとして集計し、平均をとったものである。

「該当文献中で一番多くの文献が掲載されている雑誌名を応答する機能」の有効性は薄いと評価された。認識率が以前と変わらないという結果になった理由は、先述の通り条件入力状態からなかなか移行できなかったことに起

因すると考えられる。また、認識結果を文字で逐一表示する機能については、被験者全員が非常に有用と回答した。これは被験者が全員本システム初心者であったことを差し引いても、ユーザが自分の発声をシステムがどのように理解したかを確認できることで安心するということをよく表しているといえる。

全体を通して今回の機能拡張は、ユーザに好意的に受け入れられたと見ることができる。

最後に被験者から寄せられたコメントについて考察する。

音声入力 「検索語の操作に関する入力を工夫して欲しい」という意見が多数あった。「1回の入力を短くして対話の回数を増やし、番号入力などの比較的安定した入力語で入力すれば良い」という対処案もあった。その他「聞き取りにくかった応答音声を再出力するようなコマンドを受理できれば良いのでは」という意見もあった。

表示画面 「システムの内部状態がユーザ側から見えにくいために、これに関する情報、例えばその時点で受理可能なコマンド例や、現在までに実行したコマンドの履歴のリストなどを、表示してはどうか」という意見が複数あった。

諸機能 「関連語リストからの検索を柔軟にして欲しい」「検索文字列操作のコマンドを増やして欲しい」といった意見が複数見受けられた。また、「最新文献の結果に対して、それを詳しく表示して下さい、という入力を受け付けられるようにする」といった、より知的な処理の導入を望む声も聞かれた。

応答音声 意味処理の観点から「その文献が他の何件の文献に引用されているかの情報を応答して欲しい」という意見があった。また先のシステム内部状態の不透明さとも関連するが、「ユーザの発話を黙って待ち続けるのではなく、適切な時間間隔で入力を促して欲しい」といった意見も複数あった。

システム全体 「表示画面を音声インタフェースの補助機能としてさらに有効に利用できるようにする」「画面に対してマウスやキーボードからの入力を許し、全体としての操作性を向上させるとよいのでは」といった他のメディアとの連携の強化を求める意見が複数見られた。

5 おわりに

本稿では、かねてより開発中の文献検索タスク音声対話システムに施した機能拡張を説明し、システムの評価実験の結果を述べた。

今後は、評価によって得られた被験者からの意見をもとにさらなる改良をシステムに加えていく。

また応答生成に関して、今回は応答音声の韻律情報の不具合を修正するに留まったが、今後応答生成機構の拡張に柔軟に対応していくことを睨んで韻律情報辞書の構成をより汎用的なものにすると同時に、より効果的な強調表現の生成手法を模索していくことを考えている。

さらにより高度な意味処理の導入の観点から、文献のアブストラクトの情報を活用して文献の内容を端的に応答したり、複数の文献間の意味内容的な違いを説明したりできるシステムへと拡張していくことも考えている。

謝辞

Julian を御提供頂きました、京都大学の李晃伸氏に感謝致します。

参考文献

- [1] 桐山伸也, 広瀬啓吉, “文献検索をタスクとした音声対話システムの応答生成,” 情処研報, 99-SLP-27-16, (1999).
- [2] S.Young, et al. “The HTK Book,” v2.1, Cambridge University, (1997).
- [3] 河原達也 他, “日本語ディクテーション基本ソフトウェア (98年度版) の性能評価,” 情処研報, 99-SLP-26-6, (1999).
- [4] 李晃伸 他, “文法カテゴリ対制約を用いた A* 探索に基づく大語彙連続音声認識パーザ,” 情処研報, 98-SLP-24-15, (1998).
- [5] <http://els.nacsis.ac.jp/>