

自由発話音声入力による携帯電話向けテキスト検索システム

池田 崇博 石川 晋也 三木 清一 安達 史博 磯谷 亮輔 磯 健一 奥村 明俊
NEC メディア情報研究所

{t-ikeda@di, s-ishikawa@dg, k-miki@bq, f-adachi@aj, r-isotani@bp, k-iso@bx, a-okumura@bx}.jp.nec.com

1. はじめに

近年、Web ブラウザの機能を有する携帯電話が広く普及し、外出先からでもニュース・天気予報・交通情報のような情報を容易に閲覧できるようになった。しかしながら、マニュアルやガイドブックのような大規模なテキストから必要な情報を取り出すためには、一般に、キーワード入力による検索や多階層のメニュー選択を行う必要があり、目的の情報を得るまでに煩雑なキー操作を行わなければいけないという課題があった。

そこで、我々は、ユーザが携帯電話を使って自由な言葉で検索要求を入力することが可能な音声検索システムを構築した。本システムは、携帯電話のマニュアルを検索対象とし、ユーザが携帯電話を使って入力する自由発話音声を音声認識して、その入力を検索要求としてテキスト検索を行い、検索結果を携帯電話の画面上に出力する。一般に、この種の音声検索システムでは、検索結果のユーザへの提供方法が問題となるが、本稿のシステムでは、電話回線経由で入力される検索要求を音声認識する音声サーバと、Web 経由での検索結果要求に応答する Web サーバとを連動させることで、携帯電話画面への検索結果出力を実現した[1]。本稿では、この音声検索システムに関して、特にテキスト検索手法に重点を置いて述べる。

従来、電話向けの音声入力システムとしては、IVR システムによる音声ポータルサイト等の実現例があるが、IVR システムは、ユーザの音声入力を基に、ユーザに適切な音声情報を提供するものであり、大規模テキストの検索結果を提供するものは実現されていない。

音声による入力を基にマニュアルを検索するシステムとしては、機器操作マニュアルの検索を行う対話システムが既に開発されている[2]。しかしながら、このシステムは、ビデオや FAX のような画面を持たない機器において、ユーザが、はい・いいえで返答することにより逐次的に検索結果を絞り込む手法に主眼を置くもので、本稿の検索システムとは、目的が異なっている。

一方、音声入力によるテキスト検索手法としては、TREC のクエリセットに対する研究がいくつか行われている[3][4]。しかしながら、TREC のテキスト検索が、比較的長い検索要求文を基にテキストを検索することを目的としているのに対し、本稿のシステムは、電話を通して一息で発話される程度の、比較的短い検索要

求文を基に検索を行うことを目的としており、検索の性質が異なっている。

以下では、まず音声によるマニュアル検索にテキスト検索の手法を適用する際の問題点と、本稿で提案する解決手法について述べ、その後、我々が構築したプロトタイプシステムについて説明する。また、最後にプロトタイプシステムの評価結果を示す。

2. マニュアル検索向けのテキスト検索手法

2.1. マニュアル検索の問題点

一般的に、機器のマニュアルでは、その機器の機能が網羅的に説明されており、互いに関連する多数の機能の説明が単一のマニュアル中に含まれている。この結果、類似する表現がまったく別の機能の説明に出現することもしばしばある。

マニュアルを検索する目的で発話される検索要求文を事前に収集し、分析した結果、音声によるマニュアル検索には、主として、以下の3つの問題点が存在することが明らかになった。

(1) 単一の単語を基に目的のマニュアルの項目を特定することは困難

例えば、「メール」という単語は、メールの送信、メールの受信、メールの作成等、メールに関するさまざまな機能の説明に使われるため、単体では、マニュアルの項目を特定することはできない。

(2) 否定表現と肯定表現とを区別しなければいけないケースが存在

同じ単語の組み合わせからなる文でも、否定表現と肯定表現の違いにより、別の機能を説明する場合がある。マニュアル中では、「番号を通知する」と「番号を通知しない」のように、ある設定を有効・無効にする機能の組として、しばしば出現する。

(3) 検索要求文にマニュアルにない単語が出現

一般に、マニュアル中では、機器の部位・機能等を表す用語が統一されているのに対し、ユーザは、それとは異なる単語や略語等を用いることがある。

2.2. マニュアル検索向けのテキスト検索の拡張

本稿のシステムでは、基本的に、自立語をキーワードとして、Okapi BM25 式[5]による検索モデルに基づ

いてマニュアルの検索を行う。今、検索対象となるマニュアルの各項目をパッセージと呼ぶことにすると、この検索モデルでは、検索要求文 Q に対するパッセージ P の重み W は、以下の式で表される。

$$W = \sum_{T \in Q} TW(T)$$

$$TW(T) = w \cdot \frac{(k_1 + 1) \cdot tf}{k_1 \cdot K + tf} \cdot \frac{(k_2 + 1) \cdot qtf}{k_2 + qtf}$$

$$w = \log \frac{N - n + 0.5}{n + 0.5}$$

$$K = (1 - b) + b \cdot \frac{PL}{AVPL}$$

ここで、 T は検索要求文 Q 中のキーワードを、 N は全文テキスト中のパッセージ数を、 n は単語 T を含むパッセージ数を、 tf はパッセージ P における単語 T の出現頻度を、 qtf は検索要求文 Q 中の単語 T の出現頻度を、 PL はパッセージ P の単語数を、 $AVPL$ は全パッセージの単語数の平均を表す。 k_1 、 k_2 、および、 b は予め定める定数である。

本稿では、前節で述べた 3 つの問題点に対応するため、上記検索モデルを、以下の 3 点において拡張する。

(1) 係り受け関係にある単語ペアの利用

係り受け関係にある 2 つの単語を組にして、単語ペアとして扱い、検索要求文中と同一の単語ペアを含むパッセージの重みを大きくする。具体的には、上述の W に代わり、これを単純に拡張した下記の W_{wp} を重みとして利用する。

$$W_{wp} = k_{wp}^{NP} \cdot W$$

ここで、 NP はパッセージ P と検索要求文 Q の両方に出現する単語ペアの数を表す。 k_{wp} は予め定める定数である。

これにより、例えば、「メールを送信する」という検索要求文に対しては、「メール」や「送信」という単語を単に含むだけでなく、「メール」と「送信」という単語間に係り受け関係があるパッセージをより上位にランクすることができる。

(2) 付属語による肯定と否定の区別

各単語が肯定的に使用されているか否定的に使用されているかをその単語に付随する付属語によって判別し、その単語の属性として予め付与しておく。肯定・否定の属性が一致する場合の重みを、一致しない場合の重みよりも大きくすることで、検索要求文と各単語の肯定・否定の属性が一致するパッセージをより上位にランクする。具体的には、上述の W に代わり、これを拡張した下記の W_{anc} を重みとして利用する。

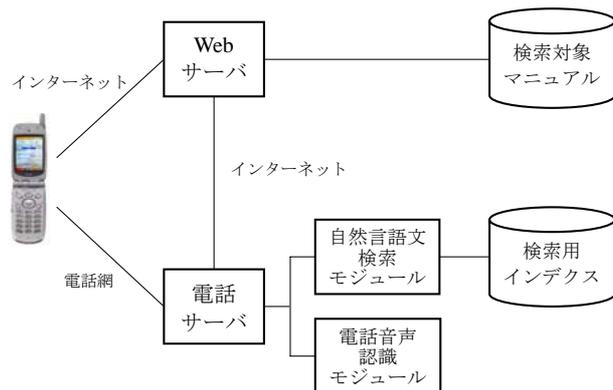


図 1: プロトタイプシステムの構成

$$W_{anc} = \sum_{T^+ \in Q} (TW(T^+) + k_{anc} \cdot TW(T^-))$$

$$+ \sum_{T^- \in Q} (TW(T^-) + k_{anc} \cdot TW(T^+))$$

ここで、 T^+ は肯定属性を持つ単語 T を、 T^- は否定属性を持つ単語 T を表す。 k_{anc} は予め定める定数である。

(3) 同義語の同一視

予め同義語辞書を用意し、同義語にあたる単語をすべて同一視して重みを計算する。具体的には、同一視したい単語が同一の同義語グループに属するように同義語グループを構成し、各同義語グループに 1 つ代表語を定めた同義語辞書を予め作成しておく。この辞書を用いて、同義語グループに属する各単語を代表語に置換してから検索処理を行う。

3. プロトタイプシステム

携帯電話マニュアルを検索対象として、前節で述べた検索手法を実装したプロトタイプシステムを構築した。検索対象のマニュアルは、約 600 ページ、約 15,000 文で構成されており、これを見出しごとに 670 パッセージに分割して使用した。

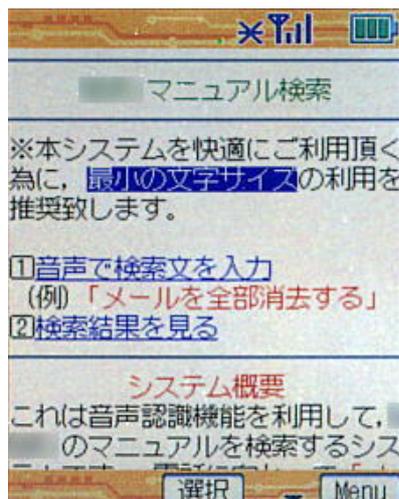
3.1. システム構成

図 1 にプロトタイプシステムの構成を示す。

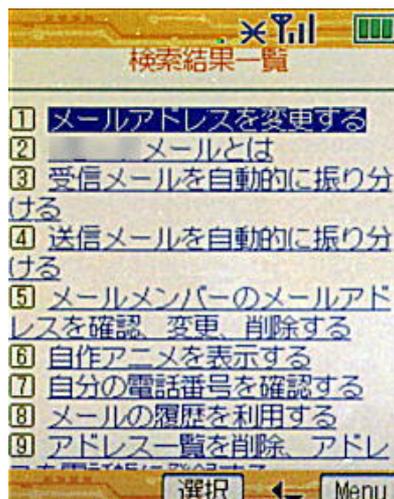
音声サーバは、ユーザからの電話を受け、電話音声認識モジュールを呼び出して、ユーザの発話を音声認識する。さらに自然言語検索モジュールを呼び出し、音声認識結果を検索要求文としてマニュアルの検索を行う。検索終了後、検索結果のリストをインターネット経由で Web サーバに送り、電話を切る。

Web サーバは、ユーザからの検索結果の要求に対して、検索結果のパッセージの見出しのリストを出力する。また、ユーザが選択したパッセージに対して、対応するマニュアル本文を出力する。

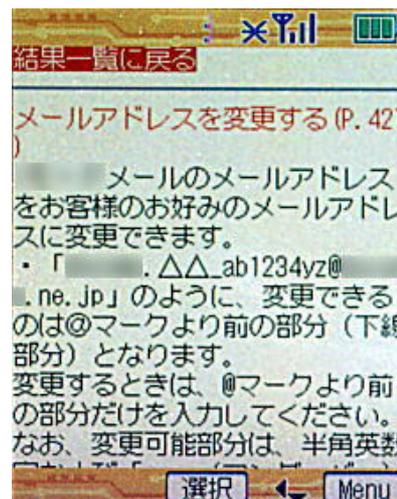
なお、本プロトタイプシステムでは、最近の携帯電話



(1) トップページ



(2) 検索結果のリスト



(3) マニュアル本文

図 2: プロトタイプシステムによるマニュアルの検索例

の画面サイズ(30 文字×15 行程度を想定)でストレスなく見ることのできる量として、上位 10 件のパッセージのみを検索結果として扱うようにしている。

3.2. 電話音声認識モジュール

電話音声認識モジュールは、電話回線経由で入力された発話を、大語彙連続音声認識により認識する。

(1) 音響モデル

学習用音声データとして、1,385 名の話者による電話発声音声データ約 20 万文を用い、性別依存の電話音声認識用音響モデルを作成した。作成したモデルは triphone で、音響分析条件はサンプリング周波数 8kHz、フレーム周期 10msec である。

(2) 言語モデル

一息で発声できる程度の長さの携帯電話に関する質問文を 20 名の被験者から約 15,000 文収集し、これを学習用テキストデータとして単語 N-gram モデルを構築した。また、携帯電話マニュアルに出現する全ての名詞を認識辞書に追加した。構築した言語モデルの 2-gram エントリは約 2 万、3-gram エントリは約 4 万、語彙サイズは約 4,000 である。

(3) デコーダ

まず、triphone 音響モデル・2-gram 言語モデルを用いて、入力音声に対してフレーム同期にビームサーチを行い、ワードグラフを生成する。次に、ワードグラフを 3-gram 言語モデルでサーチすることで、最適な認識単語列を得る。また、男性話者用音響モデルと女性話者用音響モデルを同時に用いて音声認識処理を行い、精度の高い認識結果を得ている。

3.3. 自然言語文検索モジュール

自然言語文検索モジュールは、ユーザによる発話

の音声認識結果を検索要求文として、マニュアルの検索を行う。

プロトタイプシステムでは、2.2節で述べたすべての拡張を自然言語文検索モジュールに実装した。単語間の係り受け解析は、高速化のために、構文解析ではなく、文節の係り属性と受け属性の対応を取る手法を採用した[6]。また、各定数については、前述の言語モデルの学習に用いた質問文による予備実験の結果、以下のように決定した。

$$k_1 = 100, \quad k_2 = 1000, \quad b = 0.3, \\ k_{wp} = 1.3, \quad k_{anc} = 0.3$$

同義語辞書は、言語モデルの学習に用いた質問文を基に、約 500 の同義語グループからなる辞書を構築して利用した。

3.4. システムの動作例

本システムを利用して、ユーザがマニュアルを検索する操作例を以下に示す。

- Step 1:** まず、携帯電話で本システムのトップページにアクセスする(図 2 (1))。このページには、簡単な操作説明、検索質問例と、2 つのリンクが表示されている。
- Step 2:** トップページから「音声で検索文を入力」と表示されているリンクをたどると、phone to 機能により、電話の発信が行われ、電話サーバに接続する。
- Step 3:** 音声のガイダンスに従って、検索要求文を発声する。電話音声認識モジュールによる認識と自然言語検索モジュールによる検索が終了すると、自動的に電話が切れ、トップページに戻る。
- Step 4:** トップページから「検索結果を見る」と表示されているリンクをたどると、Web サーバに接続し、Step 3 で入力した検索要求文に対する検索結果のリストが表

示される(図 2 (2))。この図は、「メールアドレスの変え方」と発声した場合の検索結果である。

Step 5: 検索結果のリストの 1 つを選択すると、それに対応するマニュアル本文を表示することができる(図 2 (3))。検索結果に適切なものがない場合には、トップページに戻り、別な検索要求文を入力する。

4. システムの評価

システム構築に用いたデータとは別に、新規の検索要求文を作成し、プロトタイプシステムを評価した。

新たに作成した検索要求文は 150 文で、これらの検索要求文に対して適合するパッセージを人手で対応付けることで、評価用のデータを構築した。実際の評価には、このうち、マニュアル中に適合するパッセージが存在した 110 文を用いた。

評価にあたっては、予め定める順位まで検索結果を見た場合に、適合するパッセージを見つけることができる検索要求文の割合を、その順位の検索成功率と定義し、1 位、5 位、10 位の検索成功率を評価尺度として利用した。例えば、110 件の検索要求文のうち、検索結果の上位 10 位以内に適合するパッセージが存在したものが 55 件あれば、10 位の検索成功率は、50%である。適合するパッセージが複数存在する場合には、そのいずれかが指定する順位までに含まれていれば、適合するパッセージを見つけることができたとして検索成功率を算出した。

まず、2.2節で述べたテキスト検索の拡張の効果を調べるため、以下の検索手法による検索成功率を比較した。結果を表 1 に示す。

BL: 拡張を行わないベースライン手法

WP: 係り受け関係にある単語ペアを使用した手法

WP+ANC: WP に加え、付属語による肯定否定の区別を行う手法

ALL: WP+ANC に加え、同義語辞書を用いた同義語の同一視を行う手法(プロトタイプシステムで用いている手法)

この結果は、どの拡張も検索成功率を向上させる効果があることを示している。特に、WP+ANC と ALL の差が大きく、同義語の同一視が大きな効果をもたらしていることが分かる。

次に、システム全体の性能を調べるために、電話音声認識結果に対する検索成功率を求めた。評価データとしては、上記の 110 文を 18 人の話者に電話を通して発声させて得られた 657 文の音声データを利用した。結果を表 2 に示す。

この結果は、80% 以上の検索要求文の発話に対して、プロトタイプシステムが検索結果として出力する上位 10 位までの検索結果を見ると適合するパッセージが含まれていることを示している。書き起こし文に対す

表 1: 検索手法ごとの検索成功率

順位	検索成功率			
	BL	WP	WP+ANC	ALL
1	40.0%	42.7%	44.5%	49.1%
5	65.5%	69.1%	70.0%	77.3%
10	73.6%	73.6%	74.5%	87.3%

表 2: 音声認識結果に対する検索成功率

順位	検索成功率
1	44.3%
5	72.5%
10	81.4%

る検索成功率(表 1 の ALL)に比べるとやや性能が劣るが、実用的に十分な性能が得られているといえる。

5. おわりに

本稿では、大語彙電話連続音声認識により、電話回線を使って音声で入力された検索要求文を受け付け、検索を行う、携帯電話向けの音声検索システムについて述べた。特に、係り受け関係にある単語ペアを利用し、付属語により肯定・否定表現を区別し、同義語を同一視することにより、テキスト検索の性能を改善した。携帯電話マニュアルを対象として開発したプロトタイプシステムでは、発話される検索要求文の 81.4% に対して、ユーザに適切な情報を提示できることが確認できた。

今後、実運用を通して、実システムの評価を行う。また、旅行ガイドや Q&A 集の検索等、別ドメインへの本システムの適用についても検討していく。

参考文献

- [1] 山田栄子 他: 2.5G 携帯電話向け音声/Web 連動型マルチモーダル UI, FIT2002, F-10, 2002.
- [2] 伊藤亮介 他: 機器操作マニュアルの知識と構造を利用した音声対話ヘルプシステム, 情報処理学会論文誌, 43, 7, pp.2147-2154, 2002.
- [3] J. Barnett et al.: Experiments in Spoken Queries for Document Retrieval, Eurospeech'97, pp.1323 - 1326, 1997.
- [4] F. Crestani: Word recognition errors and relevance feedback in spoken query processing, FQAS2000, pp.267-281, 2000.
- [5] S. E. Robertson et al.: Okapi at TREC-3, TREC-3, pp.109-126, 1995.
- [6] 佐藤研治 他: CRM 分野へ向けた日本語処理機能のミドルウェア化, 言語処理学会第 9 回年次大会, pp.109-112, 2003.