音声データベースに対する情報検索

前田勇希

平成 13 年 7 月 11 日

概 要

近年ブロードバンドネットワークの普及に伴い音声の電子化が行われつつある。しかし、現状ではよい検索手段がない。そこで本研究では、日本語ニュース記事読み上げ音声に対する情報検索手法を提案し、システムを実装してその評価を行う。本研究のシステムでは、まず音声認識器によって音素列を書き起こし、認識誤りを考慮した照合手法を用いてクエリーに合致した記事を検索する。

第1章　はじめに

1.1　研究の目的

記憶デバイスの大容量化やネットワークの広帯域化に伴い音声の電子化が行なわれはじめている。しかし、この音声を検索する手段ついてはまだよいものが無いのが現状である。そこでこの電子化された音声の検索手法について模索する。

本研究では、検索対象としてニュース記事の読み上げ音声のコーパスを用いる。ここからクエリーと同じ読みの箇所を含んだ記事の検索を目論む。音声認識を用いてコーパスの各記事に対して発音を示す記号である音素列を書き起こし、この中からクエリーの音素列と同じ読みのものを検索する。

しかし、音声認識によって書き起こされた音素列には多くの認識誤りが混入している。そのため人が聞いた場合には同一の読みであると判断される音声同士であっても同じ音素列と書き起こされるとは限らない。そこで本研究では認識誤りのモデル化を行い、これを用いて誤りに対してロバストなマッチングを試みる。これを実装しその有効性を検証する。

1.2　研究の背景と特色

1.2.1　音声データ検索に対する要求

現在、ラジオやテレビなどによってニュースや天気予報、ドラマやスポーツ中継など膨大な量の音声が放送されている。しかし、現在のところこの中からほしいものを探す有効な手段があるとはいいがたい。音声を聴くことなしに必要な情報を検索が出来るようにすることが望まれている。

1.2.2　音声認識における誤りの存在

このような音声データに対して音声認識を用いて書き起こしを生成して情報検索を行う研究が多くなされている。[3][4][5][9][1][16][6]

しかし、現在の音声認識ではかなりの率で認識誤りが伴っている。[15][14] (論文を参照) 増え続ける音声データに対して手で誤りを修正していくことは非常に大きなコストが必要である。そこで、認識結果に誤りが含まれているという前提のもとでの情報検索の手法について考える必要がある。

1.2.3 先行研究と本研究の特色

ここでは、音声認識で得られた認識誤りを含んだ音素列からの検索に関する先行研究について述べる。

RIFCDP 法

RWCP の岡らによって RIFCDP 法による、音声検索が試みられた。[6] これは、ある音声の音素パターンについて一定比率以下の誤りを許容したマッチングを行うものである。しかし、残念ながら検索の精度等に関する評価がなされておらず、これがどの程度有効であるか明らかではない。

confusion matrix

IBM の Sarivitha Srinivasan らによって Confusion Matrix を用いた音声情報検索に関する研究 [10] が行われた。これは、より高い再現率を実現するために、音素に着目した。各音素ごとにどういう音素へと誤りやすいかという統計を用意し、これを用いてある音素列がどのような音素列と誤りやすいか推定し、認識誤しやすい音素列も含めて検索する。ビデオ検索について評価が行われ、特に音声認識の辞書にない語について有効であった。

OCR 文書の検索検索

その他、音声検索に似たタスクとして、OCR 文書の検索というタスクがある。OCR によって得られた認識誤りを含んだ文書からの情報検索である。[2][11] 東京都立大の太田は、英文 OCR 文書からの誤りに対してロバストな検索手法として confusion Matrix の拡張手法 (ECMR) を提案している。[8] これは、時間軸上で前にどの文字が出現したかを考慮した Confusion matrix を用いて英文 OCR の認識誤りをモデル化し、このモデルに基づいて OCR 文書のロバストな検索を実現している。

本研究の位置づけ

本研究では、音素列キーワード[1] を含んだ 100 秒前後の長さのニュース記事音声からの検索を試みる。RIFCDP を単純化した連続 DP マッチング法をベースとして Confusion matrix を組み込み、認識誤りに対してロバストな検索の実現を目指す。

[1] 音声認識器の辞書にある語ではなく任意の語をキーワードとする

第2章　音声認識

2.1　音声認識器

本研究では、音声認識器として日本語ディクテーションシステム julius [7] [12] を利用することとした。これは、大語彙連続音声認識研究開発の共通のプラットホームとして開発設計された。このプラットフォームは、標準的な認識エンジン、日本語音響モデル、日本語言語モデルおよび日本語形態素解析、読み付加ツール等から構成されている。日本語の音声に対してかな漢字交じり文を書き起こすことができる。

2.2　音素の書き起こし

ニュース音声に対して、かな漢字交じり文の書き起こしを行うためにはニュースで用いられている語彙が音声認識器の辞書に含まれていなければならず、 また同音異義語などや形態素区切りを適切に処理する必要がある。しかし、 ニュースでは常に新しい語が出現する。たとえば、ある国で大統領が変わる度に新しい 固有名詞が誕生する。

本研究では、通常のかな漢字交じり文ではなく、発音を示す記号である「音素」を対象とした検索を試みる。

音素認識には辞書が必要ないため、語彙の制限のない検索が可能である。

julius では、特殊な辞書ファイルを用意することによって言語モデル部[1] を切り放し音響モデル音響情報から音素列を出力するモジュールの出力する音素だけを得ることが可能である。表 2.1 に julius が出力する音素を示す。本研究ではこの julius 音響モデルの出力する音素を対象として検索を試みる。

a i u e o a: i: u: e: o: N(ん)
w y p py t k ky b by d dy g gy
ts ch m my n(な行) ny h hy f s
sh z j r ry q(っ)

図 2.1: julius が出力する音素

[1] 音素列からかな漢字文を出力するモジュール

第3章　コーパス

3.1　コーパスの特徴

研究のための素材としてRWCPによって整備されたニュース音声コーパスを用いることとした。ノイズのない環境で録音されたものであるため、様々なノイズを含んだ実際の放送ニュースと比べて現在の音声認識器でも比較的高い精度で認識することができる。

3.2　コーパスの内容

全246記事、6人(女性3人、男性3人)のアナウサーによる音声である。各記事は100秒前後の長さで1000個前後の音素から構成される。このコーパスには、音声の他に図3.1のような人手で作成されたかな漢字交じり文の書き起こしと読み仮名が付属している。

3.2.1　書き起こし

この音声コーパスについて、音声認識器juliusで音素列の書き起こしを生成した。これを図3.2に示す。耳で音声ファイルを聞いた結果とは異なっており、多く音声認識誤りを含んでいることが確認された。

0001
500
5200
会計帳簿の紛失で、巨額な活動資金の詳細が不明だった
かいけえちょおぼのふんしつできょがくなかつどおしきんのしょおさいがふめえだった

図 3.1: コーパスに附属するテキスト

k a i k i: ch o: b o n o f u i q s u d e sp ky o N u k u n a k a
ts d o: sh k i n o: sh o: s a N g a f u m e: d a q t a sp n a u: n
o t o: k gy o n i N p i k u sh o: ch i i N t a i d e: s r n a q sp
k o u · · ·

図 3.2: juliusの出力した音素列の例

第4章　音声認識誤り

4.1　音声認識誤りについて

4.1.1　はじめに

音声認識誤りとはどのようなものであろうか。まず、音素に着目してニュース音声の読み上げとその音声認識による書き起こしのプロセスを通信システムとして一般化して捉えた場合に認識誤りはどのような形でとらえることができるかを述べる。

4.1.2　記事の読み上げと音声認識

人がニュース音声を読み上げ、それをコンピュータが音声認識で書き起こすプロセスについて、語の読みである音素に着目して通信システムになぞらえると以下のとおりである。

1. 記事中の語から音素を想起する (情報源)
2. 音素を発声する (符合化)
3. 音声が伝搬 (通信路)
4. マイクで音声を収集し、音素を認識 (復号化)
5. 音素から語を書き起こす (あて先)

1. の情報源とは、情報を発生する源である。ここでは、ニュース記事中の各語を読み取り語の音素を想起するまでに相当する。2. の符号化とは、通報が通信経路を通過出来るように変換する作業である。音素はそのままでは空気中を伝搬することは出来ないため、音素を声帯で発声し音声とすることである。3. の通信路とは音声が空気中を伝搬することに相当する。4. の復号化は、音声から音素を認識することに相当する。

一般に音声認識誤りとは、1. の人間が発声した語と 5. の認識された語の間になんらかの理由で解離がみられることである。

本研究では、特に 1. の音素と 5. の音素が異なっているときに音声認識誤りが発生したとする。すなわち、符号化や通信路、復号化の過程でなんらかの誤りが混入し、情報源の音素とあて先に到達した音素が異なったものとなることである。

符号化、通信路、復号化については、本研究ではブラックボックスとして扱う。符号化の方法、すなわち発声の問題や音響的な問題、音声認識の問題については対象とはしない。そのかわり、全体をとおしてどのような誤りが混入するかをみる。

k a i k e e ch o o b o n o f u N sh i ts u d e ky o g
a k u n a k a ts u d o o sh i k i N n o sh o o s a i
g a f u m e e d a q t a ···

図 4.1: 読み仮名から音素を生成

k a i k i i ch o o b o n o f u i q s u d e
ky o N u k u n a k a ts d o o sh k i n o o
sh o o s a N g a f u m e e d a q t a ···

図 4.2: 音素列を変換した例

4.2 認識誤りの調査

音声認識の際にどれくらいうまく認識でき、どれくらい認識誤りを起こすかを調査する。音素列書き起こしの比較と認識結果の集計について述べる。

4.2.1 音素列の比較

ニュース音声コーパスに対して音声認識を実行し、音素列を書き起こした。図 3.2 にその一部を示す。この音素がどの程度誤りを含んでおり、どの程度信頼できるかを調査する。そこで、比較のために音声コーパスに付属している人手で作成されたテキストを用いることとした。(図 3.1) ここに含まれている読み仮名を抽出し音素列を自動生成するプログラムを作成した。(図 4.1)

しかし、この音素列の中ではすべての長母音[1] が二重母音[2] として表現されている。そのため、図 4.2 のように音声認識で得られた音素列中の長母音を二重母音へと変換し、比較可能な形に変換するプログラムを作成した。

4.2.2 比較プログラム

音素列同士を比較し、三種類の誤り「置換、欠落、挿入」を検出しログとして出力するプログラムを作成した。置換のあやまりは、「::x::a:c:b:」。すなわち、音声認識結果の文脈 axb において本来は x であるべきものが認識結果では y へと置き換わっていると表現する。欠落の誤りは、仮想文字 ϕ を用いて「::x::a:phi:b:」とする。文脈 ab において、a と b の間に入るべき x が欠落しているとする。挿入の誤りは、仮想文字 ϕ を用いて「:: phi::a:x:b:」とする。文脈 axb において x が不要であるにもかかわらず余計に挿入されているとする。

本プログラムでは、DP マッチングを用いてこれらの誤りの数を最小化する。なお、置換は欠落と挿入の連続したものとして表現することが出来る。ここで判断の曖昧性が発生する。そこで、誤りの個数を最小化するため置換の重みを 3、挿入及び欠落を 2 とし、この重みの合計

[1] 「カー (k a:)」の「a:」のように長い母音
[2] 母音二つが連続したもの。「aa」など

	正解 認識結果
	: k : a : : k : a
	k : a : i : k : a : i
	a : i : k : a : i : k
	i : k : e : i : k : i
	k : e : ch : k : i : ch
すなわち、	e : e : ch : i : i : ch
	e : ch : o : i : ch : o
	ch : o : o : ch : o : o
	o : o : b : o : o : b
	o : b : o : o : b : o
	b : o : n : b : o : n
	o : n : o : o : n : o

図 4.3: 音素の比較ログ

認識結果 : 実際の音素 : 確率
k o o : o : 0.79482072
k o o : p : 0.00000000
k o o : q : 0.00000000
k o o : r : 0.00000000
k o o : s : 0.00000000
k o o : t : 0.00000000
k o o : u : 0.06374502
k o o : w : 0.00000000

図 4.4: 誤りの統計

値を最小化するようにすることとした。

その結果、置換とみなすことが出来る欠落挿入をすべて置換とすることが出来、誤りの総数の最小化を実現した。プログラムの出力の一部を図 4.3 に示す。比較結果によると、音素認識の誤り率はおよそ 34%であった。

4.2.3 集計

誤りの比較結果を元に、どのような誤りが起りやすいか、集計をとった。ある音声認識結果に対して、実際にはどの音素であるかをまとめた。(図をかこう、そのうち)

第5章　ロバストなマッチング

5.1　処理の目的

クエリー音声に適合した、ニュース記事の朗読音声を選びだすことが目的である。そのため、クエリー音声とニュース記事音声の間に関連があるか否かを知る必要がある。

5.2　処理の概要

ニュース記事の朗読音声から音声認識器によって音素列を書き起こす。また、検索のためのクエリーは音素列として与えらる。記事の朗読音声から音声認識で生成した音素列とクエリーの音素列の距離を計算し、関連があるか否かを判定する。

しかし、記事の音声認識の際にはしばしば認識誤りが混入する。人間の耳で聞き取ることが出来るものと同じように認識できるとは限らない。音素が欠落する、あるいは逆に余計な音素が挿入される、異なる音素に置き換わる、などの誤りがしばしばみられる。誤りなく認識された音素列だけでなく、認識誤りににによって長さが変動した音素列も検索できるようにすることが望ましい。

そこで、記事の音素列中の任意の始点からはじまる任意の長さの音素列についてクエリーの音素との距離を測ることとする。この距離が近いものを合致したものとする検索アルゴリズムについて検討する。

第 3 節で音素列同士の比較のためにまず音素同士の距離の算出方法について考え、つづく第 4 節でこれを元にした音素列間の距離の算出方法、第 5 節で距離に応じた判定方法について述べる。

5.3　音素間の距離計算手法

ここでは、本研究で評価を試みる 2 つの距離計算方法、exact matching, confusion matrix について述べる。

5.3.1　exact matchingaによる距離計算

厳密な一致による距離計算である。二つの音素が同一のものであるとき距離がゼロであるとし、同一ではないときに距離が一であるとする。

5.3.2 confusion matrix による距離計算

音素が同一か否かによってゼロか一かの二者択一をするのではなく、似ている音素の場合にはゼロよりも大きな値を与え、似ていない音素の場合はゼロに近い小さな値を与える方法について述べる。最初に基本となる考え方について説明した後、続いて confusion matrix とこれを求める方法について述べる。最後に音素間の距離の計算方法について述べる。

記憶のない通信路

音声認識誤りは、さまざまな要因から発生する。その要因について述べることは本研究の範疇を越えるためここでは述べない。しかし、その影響は、通信路 (ここでは、符号化、復号化も含める) への入力として与えられた音素ごとに変わっていることが観測できる。すなわち、ある音素 t が発声され音声として伝搬し、それが音声認識されて音素 r として書き起こされるとき、音素 t がどの音素であるかによって、音素 r のとりうる確率分布が異なるということである。例を述べると、音素"ny" (にゃ "ny a" の子音) を発声して認識する場合、音素"n"(な "n a "の子音) へと書き起こされる確率は比較的高い。しかし、音素 ny(あ "a") が音素 k(か "k a " の子音) へと書き起こされる確率は低い。

ここで、出力の音素の確率分布は、入力として与えられる音素以外からはいかなる影響も受けないものとする。すると、出力の確率分布は入力として与えられる音素によって決定されることとなる。

このような通信路を記憶のない通信路と呼ぶ。本研究では、発声から音声伝搬、音声認識の仮定までを記憶のない通信路とみなすこととする。

confusion matrix

Confusion matrix[3][10] というものが提案されている。

これは、任意の音素 r が音声認識結果中に観測されたときにそれが実際には任意の音素 t である確率の推定値を $C(t, r_n)$ という行列へとまとめたものである。通信路に入力音素 t が与えられたとき、出力として r_n が得られることに相当する。

この行列の値は、音声認識によって生成された音素列と人手で作成された読み仮名から生成された音素列を比較した結果から求めることが出来る。ニュース記事中の訓練用正解データ中にある音素 t が出現するとき、この音素 t に対応して音声認識からの書き起こしの中に観測される音素 r_n を求める。対応する音素が欠落して観測されない場合には空音素 ϕ を割り当てる。音素 t に対応する音素 r_n の出現回数を、音素の種類ごとにカウントしていく。そして、音素 r_n の出現回数を音素 t の総出現回数で割ることによって、音素 t が生起するときにある音素 r_n が生起する確率 p_n の推定値が得られる。この値を以下の式で confusion matrix に格納する。

$$C(t, r_n) = p_n \tag{5.1}$$

なお、音素の種類を k 個とすると、任意の音素 t について式 5.3 が成り立つ。

$$C(t, r_1) + C(t, r_2) \cdots C(t, r_k) = 1 \tag{5.2}$$

格納した結果の一部を表 5.1 に示す。

表 5.1: 1 次元 Confusion Matrix の一部

t	r	C(t,r)
i	g	0.00074710
i	h	0.00049807
i	py	0.00037355
i	i	0.88930395
i	j	0.00062259
i	ry	0.00062259
i	ch	0.00000000
i	k	0.00149421
i	m	0.00000000

confusion matrix の拡張

さて、音素 r_n が一音素である場合について述べた。ここで、r_n をスカラーではなく音素のベクトルであるとする考えについて述べる。ある音素 t が訓練用データ中に出現するとき、それに対応した音声認識結果中に観測される音素 r_{n1} とその前あるいは後ろに出現する音素 r_{n2} が同時に出現する回数をカウントする。これを 1 次の confusion matrix と同様に音素 t の総出現回数で割ることによって、音素 t が出現するときに音素 $r_{n1}r_{n2}$ が出現する確率を求める。これを 2 次元 confusion matrix とよぶ。

r_n を 3 次元の音素ベクトルとする場合についても同様に計算することが出来、このような confusion matrix を 3 次元 confusion matrix とよぶ。

また、r_n がスカラーである場合の confusion matrix は、これを 2 次元 confusion matrix などと区別するために 1 次元 confusion matrix とよぶ。

なお、confusion matrix はすべての音素 t と音素 r についてゼロより大きい値が求められるわけではなく、音素 t と音素 r のペアによっては、出現しないため値がゼロとなる場合がある。今回 confusion matrix の作成に用いたデータでは、1 次元の場合は約 40%がゼロであり、2 次元では約 90%, 3 次元では約 99%がゼロであった。

音素認識誤り確率

ある音素 t と音素 r についての音素認識誤り確率を、以下の式で定義する。

$$P_{Y_r|X_t}(y_r|x_t) = C(t,r) \tag{5.3}$$

音素間距離計算

音素 t と音声認識誤りを含む音素 r の間の距離 $d(t,r)$ を、音素認識誤り確率を用いて以下の式で定義する。

$$d(t,r) = 1 - P_{Y_r|X_t}(y_r|x_t) \tag{5.4}$$

5.4 音素列間の距離計算手法

続いて、クエリー音素列と記事音素列の間の距離計算手法について検討する。クエリー音素列と、記事音素列中の任意の場所に含まれているクエリーに似ている音素列を比較し、距離を計算する手法に付いて述べる。

5.4.1 連続 DP マッチング

連続 DP マッチング法を用いた距離計算手法について述べる。

クエリーとして与えられた音素列の長さを J とし、比較対象の記事の音素列長さを I とする。まず、連続 DP マッチング法で計算のバッファとして用いる行列 $g(I,J)$ と、行列 $g(I,J)$ を音素列長さについて正規化するための行列 $c(I,J)$ を用意する。各行列は、からかじめゼロで初期化しておく。続いてここから式 5.6 と式 5.7 に基づいて $i=1,j=1$ から $i=I,j=J$ まで再帰的に値を計算する。

$$g(i,j) = min \begin{cases} g(i-2)+2d(i-1,j)+d(i,j) & (a) \\ g(i-1,i-1)+2d(i,j) & (b) \\ g(i-1,j-2)+2d(i,j-1)+d(i,j) & (c) \end{cases} \tag{5.5}$$

$$c(i,j) = \begin{cases} c(i-2,i-1)+3 & \text{if (a)} \\ c(i-1,j-1)+2 & \text{if (b)} \\ c(i-1,j-2)+3 & \text{if (c)} \end{cases} \tag{5.6}$$

なお、$d(i,j)$ は音素 i と音素 j の間の距離である。

記事の音素列中の任意の始点から始まる音素列 (長さは 1/2J から 2J の範囲で任意) と、クエリー音素列との間の距離 $D(i)$ を以下の式 5.8 で定義する。Di の最小値を記事とクエリーの距離 D とする。

$$D(i) = g(i,J)/c(i,J) \tag{5.7}$$

5.4.2 傾斜制限無し連続 DP マッチング

これまでのべてきた連続 DP マッチングでは、記事の音素列は 1/2J から 2J の範囲内であるという前提に基づいている。音声認識誤りが多く混入する場合にはこの範囲におさまらないことが考えられる。そこで、この傾斜制限を撤廃したものについても考える。まず、DP マッチングのためのバッファ $g(I,J)$ を以下のように初期化する。

$$g(i,j) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} \tag{5.8}$$

続いて $g(I,J)$ の値を以下の式に基づいて $i=1,j=1$ から $i=I,j=I$ まで順に再帰的に計算する。

$$g(i,j) = max \begin{cases} g(i-1,j)+d(rphone_i,\phi) \\ g(i-1,j-1)+d(rphone_i,tphone_j) \\ g(i,j-1)+d(\phi,tphone_j) \end{cases} \tag{5.9}$$

記事と音素の間の距離を以下に式に基づいて定義する。

$$D(i) = g(i, J)/J \tag{5.10}$$

この $D(i)$ のうち最小のものをクエリーと記事の距離 D とする。

5.5 判別方法

クエリーの音素列と記事の音素列の距離の計算方法について述べた。ここでは、この距離を元にクエリーと記事が関連あるものかどうかを判別する方法について述べる。

5.5.1 音素列長比例判定

D は、クエリーの音素数にほぼ比例して値が大きくなる。単純に D の値で判別するだけでは、クエリーの音素数による影響を受けて正しく判定できない。そこで、以下の式 5.12 で判定を行う。なお、α はマッチングの挙動を制御するための係数であり、q_n はクエリーの音素数である。

$$D > \alpha \cdot q_n \tag{5.11}$$

5.5.2 音素生起確率による判定

実験中。

第6章　検索システムの実装

6.1　実装の目的

これまで述べてきた検索手法の有効性を検証するため、これを実装しニュース音声の検索について評価を試みる。

6.2　システムの構成

日本語音声全文検索エンジン「じゃいサーチ」を実装した。構成を以下に示す。

1. サーバ部 (web サーバ上で CGI として動作)
   (a) 検索クエリー受付部
      - 読み仮名としてクエリーを受理し、音素列を出力
   (b) データベース検索部
      - クエリー音素列とニュース記事音素列とのマッチングを実行し、距離が閾値を下回るものを関連した記事と判定
   (c) 出力部
      - 記事名と音声ファイルを出力
2. クライアント部
   - web ブラウザで閲覧する

6.3　プラットホーム

第7章　評価実験

7.1　評価実験の目的

まず、検索のベースラインとして厳密なマッチングによる検索を試み、続いてよりロバストと思われるマッチング手法について試みる。

7.2　用意するクエリー

ここでは、89 記事に対して 50 個のクエリーを用意した。各クエリーは少なくとも 5 記事以上に含まれているものである。平均記事数は

7.3　訓練用データと評価用データ

全 246 個の記事から 91 個を抽出し誤りの傾向を知るための訓練用データとし、89 個を抽出して評価用のデータとした。評価用データセット対して、音素列クエリー 50 個を用意。クエリーはのべ xxx 個の記事を参照している。一クエリーあたりの参照記事数は約 xx 記事である。ここで、

- 再現率 = 検索された文書中の該当文書の数/全文書中の該当文書の数
- 精度 = 検索された文書中の該当文書の数/検索された文書数

として 50 クエリーについてそれぞれ再現率と精度を求めた。また、全体をとおしてみるため、平均再現率、平均精度を計算した。

7.4　exact matching

音声認識で得られた音素列に対して通常のテキストと同じ方法で検索を行い、再現率、精度を求めた。(表 7.1) 認識誤りを全く考慮していないため、再現率、精度ともに大きく下がっていることがわかる。

file=eva$_l og_0$.*eps*

図 7.1: 連続 DP マッチング 傾斜制限あり 0 次マルコフ

音素	文字	出現文書数	
y u ny u u	輸入	7	
k e N ky u u	研究	5	
k a b u sh i k i sh i j o o 株式市場	6		
k e e z a i	経済	39	
k a b u k a	株価	5	
h e e k i N	平均	15	
ch u u sh o o k i gy o o	中小企業	6	
g u t a i t e k i	具体的	5	
a m e r i k a	アメリカ	14	
k i s e e k a N w a	規制緩和	10	
o o s a k a	大阪	7	
hy a k u e N	百円	8	
k a k a k u	価格	12	
k a n a d a	カナダ	6	
h a N sh i N d a i sh i N s a i	阪神大震災	15	
g o g o	午後	9	
d o i ts u	ドイツ	5	
g i N k o o	銀行	25	
t o o ky o o	東京	15	
o o t e	大手	30	
hy o o k a	評価	10	
d o r u	ドル	22	
n i ch i g i N	日銀	8	
e N d a k a	円高	10	
t o o k i	投機	7	
z e N k o k u	全国	??	
h o k e N	保険	7	
k o N k a i	今回	25	
n i q p o N	日本	15	
s e e f u	政府	30	
j i d o o sh a	自動車	9	
sh o o k e N	証券	9	
o o k u r a sh o o	大蔵省	8	
j o o sh o o	上昇	9	
g e r a k u	下落	6	
t o r i h i k i	取引	11	
s a k u g e N	削減	9	
j u u gy o o i N	従業員	5	
m i N k a N	民間	10	
s e k a i	世界	9	
s a i m u	債務	5	
s a i k e N	債権 (再建)	9	
ch o o k i	長期	5	
ts u u k a	通貨 (通過)	11	
a k a j i	赤字	13	

表 7.1: 評価結果

厳密なマッチングの結果	
recall	precision
18.47%	45.95%

file=eva$_l og_1$.$eps, width = 10cm$

図 7.2: 連続 DP マッチング 傾斜制限あり 1 次マルコフ

file=eva$_l og_2$.$eps, width = 10cm$

図 7.3: 連続 DP マッチング 傾斜制限あり 2 次マルコフ

file=eva$_l og_3$.$eps, width = 10cm$

図 7.4: 連続 DP マッチング 傾斜制限あり 3 次マルコフ

file=eva$_l og_1 0$.$eps, width = 10cm$

図 7.5: 連続 DP マッチング 傾斜制限なし 0 次マルコフ

file=eva$_l og_1 1$.$eps, width = 10cm$

図 7.6: 連続 DP マッチング 傾斜制限なし 1 次マルコフ

file=eva$_l og_1 2$.$eps, width = 10cm[width = 10cm, clip] eva_l og_1 2$.$eps$

図 7.7: 連続 DP マッチング 傾斜制限なし 2 次マルコフ

file=eva$_l og_1 3$.$eps, width = 10cm$

図 7.8: 連続 DP マッチング 傾斜制限なし 3 次マルコフ

file=eva$_l og_2 1$.$eps, width = 10cm$

図 7.9: 連続 DP マッチング 傾斜制限有り NCM1 次マルコフ

file=eva$_l og_2 2$.$eps, width = 10cm$

図 7.10: 連続 DP マッチング 傾斜制限有り NCM2 次マルコフ

file=eva$_l og_2 3$.$eps, width = 10cm[width = 10cm, clip] eva_l og_p 3$.$eps$

図 7.11: 連続 DP マッチング 傾斜制限あり NCM3 次マルコフ

7.4.1 まとめ

傾斜制限を導入したほうが高い精度が得られた。(比較結果より) 傾斜制限に加えて、正規化した confusion matrix(NCM) を導入する場合がもっとも高い精度をえられた。しかし、一次近似以上については有効な結果が得られなかった。バグと思われる。

傾斜制限の導入下で、0 次近似 CDP と NCM1 次近似 CDP が高い結果であったが、後者が少し良い結果となった。

クエリー yunyuu について検索した例を示す。NCM1

y u ny u u f1n2003.phn.rwcp 0.199327 u u j u u f1n2013.phn.rwcp 0.111111 y u u u u f1n2014.phn.rwcp 0.197584 j u u u u f2n2168.phn.rwcp 0.199486 u u sh u u f4n2071.phn.rwcp 0.125000 y u ny u u m1n2031.phn.rwcp 0.125000 k u g u u m1n2033.phn.rwcp 0.125000 e m ny u u m1n2040.phn.rwcp 0.125000 t o y u u m1n2041.phn.rwcp 0.185460 ry u ny u u m2n2126.phn.rwcp 0.122101 u N ny u u m2n2141.phn.rwcp 0.165305 ny y u u fin

0 y u ny u u f1n2003.phn.rwcp 0.200000 u u j u u f1n2013.phn.rwcp 0.111111 y u u u u f1n2014.phn.rwcp 0.200000 j u u u u f2n2168.phn.rwcp 0.200000 u u sh u u f4n2071.phn.rwcp 0.125000 y u ny u u *f4n2075.phn.rwcp 0.200000 u u ny u u m1n2031.phn.rwcp 0.125000 k u g u u m1n2033.phn.rwcp 0.125000 e m ny u u m1n2040.phn.rwcp 0.125000 t o y u u m1n2041.phn.rwcp 0.200000 ry u ny u u m2n2126.phn.rwcp 0.125000 u N ny u u m2n2141.phn.rwcp 0.166667 ny y u u *m3n2103.phn.rwcp 0.200000 u u ny u u fin

NCM1 のほうが精度 (precision) が高い結果となった。生起頻度の低い音素について余計な拡張を行わないためであると考えられる。ただし、その差はごくわずかであり、有意差があるとみなしてよいか判断できなかった。

第8章　まとめ

8.1　今後の予定

関連図書

[1] BO-REN BAI, BERLIN CHEN, and HSIN-MIN WANG. Syllable-based chinese text/spoken document retrieval using text/speech queries. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(5):603–616, 2000.

[2] K.Marukawa, Tao Hu, Hiromichi Fujisawa, and Yoshihiro Shima. Document retrieval tolerating character recognition errors - evaluation and application. *Pattern Recognition*, 30(8):1361–1371, 1997.

[3] Kenney Ng. *Subword-based Approaches for Spoken Document Retrieval*. PhD thesis, Massachusetts Institute of Technology, 2000.

[4] NIST TREC. *The TREC Spoken Document Retrieval Track: A success story*, 2000.

[5] Proceedings of 8th Text Retrieval Conference. *The THISL SDR system at TREC-8*, 2000.

[6] Proceedings of ESCA Eurospeech Conference. *Referring in Long Term Speech by using Orientation Patterns Obtained from Vector Field of Spectrum*, 1997.

[7] Proceedings of IEEE workshop on Automatic Speech Recognition and Understanding. *Japanese dictation toolkit - plug-and-play framework for speech recognition R&D*, 1999.

[8] Proceedings of International Conference on Document Analysis and Recognition. *Retrieval Methods for English-Text with Miss-Recognized OCR Character*, 1997.

[9] Proceedings of International Conference on Spoken Language Processing. *Multi-scale audio indexing for chinese spoken document retrieval*, 2000.

[10] Proceedings of SIGIR2000. *Phonetic Confusion Matrix Based Spoken Document Retrieval*, 2000.

[11] Proceedings of Workshop on Information Retrieval with Oriental Language. *Probabilistic Retrieval Methods for Text with Miss-Recognized OCR Character*, 1996.

[12] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 山本幹雄, 山田篤, 宇津呂武仁, and 鹿野清宏. 日本語ディクテーション基本ソフトウェア (98 年度版). **日本音響学会誌 (技術報告)**, 56(4):255–259, 2000.

[13] 今井秀樹. **情報理論**. 昭晃堂, 1984.

[14] 鷹尾 誠一, 緒方 淳, and 有木 康雄. ニュース音声に対する検索方法の比較. **音声言語情報処理**, 29(17):97–102, 12 1999.

[15] 西崎 博光 and 中川 聖一. キーワードの音声入力によるニュース音声の検索法. **音声言語情報処理**, 29(16):91–96, 12 1999.

[16] 遠藤 隆, 張 建新, 中沢 正幸, and 岡 隆一. 音声データの自己組織化と音声検索システム. **音声言語情報処理**, 25(4):19–24, 2 1999.