ニュース音声データベースの検索システム

西崎博光　中川聖一
豊橋技術科学大学 情報工学系

1 はじめに

近年、多くのニュース番組が放映されているが、過去のニュース番組から興味のある記事を見つけたいという欲求がある [1]。

数多くのニュース番組の中から、必要なニュースを見つける場合、各ニュースに対してインデックスが付けられている場合はそれを使って検索することができる。しかし、ニュース音声データからの検索に対する需要もあり、この場合は放送された全てのニュースを予め文字化し、データベースとして蓄積しておく必要がある。この作業を人手で行なうのは不可能に近く、大語彙音声認識システムを用い、自動的に書き起こすこととなる。

そこで本研究では、自動的に書き起こしたデータベースでの検索性能を調べるため、まず、実際のニュース音声に対して、音声認識システムにより書き起こし、検索用データベースを作成した。このデータベースと正確に書き起こしたデータベースに対して、キーワード群を使って検索された記事の一致率を求め、比較した。実験の結果、単語認識率が低いにもかかわらず高い一致率が得られた。

キーワードを音声で入力することを考えた場合、必ずしも正しく認識されるとは限らない。また、機械には認識結果が正しいキーワードかどうかわからないので、誤りもありうる認識結果を使って検索を行なわざるを得ない。また、キーワードに同音意義語が存在する場合は、同じ読みの単語すべてをキーワードとして扱う必要がある。こういった場合、実際にユーザーが意図しない記事を大量に含む検索結果が得られたり、逆に全く結果が出力されないことになるので、これらの記事をうまく絞り込んでいく必要がある。そこで検索処理に先立って、単語間の関連度を用い、キーワード候補の語数を絞る手法を提案する。単語間関連度は正確に書き起こしたデータベースより学習し、キーワード候補をグルーピングする。その結果、キーワード候補は幾つかのグループに区分される。そして、単語数の最も多いグループ中の単語を用いて検索処理を行なう。

本稿では実際にキーワードを音声で入力し、前述の手法で検索実験行ない、その有効性を示す。

2 ニュース検索システムの概要

今回試作した、ニュース検索システムの概略図を図1に示す。

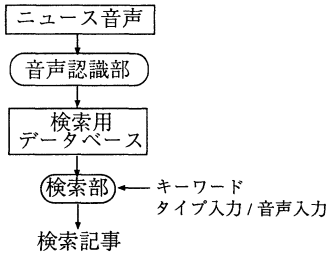

図 1: システムの概略

まず、ニュース音声を音声認識システムに通し、自動的に検索用データベースを作成する。これを基に、入力キーワードに応じて記事を検索部で検索する。

検索部では全文検索を行なっているが、インデックス法 [2] を用いることで、高速な検索を可能にしている。検索キーワードは、テキスト入力でキーワードをいくつか入力する。すべてのキーワードが完全に一致した記事のみを出力する。ただし、これでは制約がきつ過ぎ必要な記事が検索されないので、入力キーワード数が多い場合は、全部が一致しなくてもその大部分 (3分の2以上) が一致している記事を出力する。もし、入力キーワードが未知語だった (音声認識で使用した語彙辞書に入っていない) 場合は、音節列 (かな文字列) 単位の DP マッチングを行なうようにしている。

3 キーワードのテキスト入力による検索実験

前節で説明した検索システムで検索実験を行なった。実験対象の音声データは、NHK ニュース (1996年6月1日～7月14日) で、記事の数は 673 記事、文数で 3951 文である。ニュース音声の書き起こしに使用した音声認識システムの条件を表 1に示す。言語モデルは第1パスでは語彙サイズ 20000 の単語 bigram、第2パスでは単語 trigram を使用している。この音声認識システム [3] を使用した場合、ニュース音声 (背景雑音なども混入されている) の単語カバー率は 96.6%、単語正解率は 71.0%、単語正解精度は 61.7%となった。

キーワードを選択するため、1記事につき3人の被験者にキーワードを3～5個選んでもらい、3人とも共通に選んだ単語の集合をキーワード群とした。

正確に書き起こしたデータベース (以後「データーベース (A)」と記す) と、自動的に書き起こしたデータベース (以後「データーベース (B)」と記す) に対して、キーワード群を使って検索された記事の一致率を求めた。実験結果を表 2に示す。表 2で、

An Information Retrieval System of Broadcast News Speech Documents
Hiromitsu Nishizaki and Seiichi Nakagawa
Department of Information and Computer Sciences, Toyohashi University of Technology

表 1: 認識実験の実験条件

音響モデル	
5 状態 4 出力分布 (4 混合ガウス分布, 全共分散行列)	
離散継続時間分布付き連続出力分布型 HMM	
音節カテゴリ数	113 音節
サンプリング周波数	12kHz
窓関数	21.33ms ハミング窓
フレーム周期	8ms
分析	14 次元 LPC 分析
学習データ	
ASJ ATR503 文 A〜J セットの 6 名の男性話者と 216 単語の音声データから初期モデルを作成	
ASJ ATR503 文 A〜J セットの 30 名の男性話者と JNAS 新聞記事文 125 名の男性話者を MAP 推定で追加学習 (総発話数 17221 文)	
特徴パラメータ	
LPC メルケプストラム (10 次元 × 4 フレームの特徴量を KL 展開で 20 次元に圧縮)	
+ Δ ケプストラム (10 次元)	
+ ΔΔ ケプストラム (10 次元)	
+ Δ パワー + ΔΔ パワー	

一致率： あるキーワード群でデータベース (A) に対して検索された数個の記事が、同じキーワード群でデータベース (B) に対して検索した場合に、どれだけ検索されたかを表す割合。

検索率： 対象記事がどれだけ正しく検索されたかを表す割合。

湧きだし記事数： 検索された余計な記事数。

である。1 キーワード群当たり検索された記事数はデータベース (A) で平均 4.1 記事、データベース (B) で平均 4.7 記事、一致率は 86.3%であり、対象記事 30 記事中で (B) のデータベースで 27 記事 (90.0%) が正しく検索された。また、余計な記事が検索された割合は、1 記事当たり 1.2 記事と低くなっている。

表 2: テキスト入力での実験結果

検索対象記事	： 30
一致率	：86.3%
検索率	：90.0%
湧きだし記事数	： 37

表 2の実験結果を見ると、単語正解率 71.0%、単語正解精度 61.7%とかなり低い値になっているにもかかわず一致率が高くなっている。これは、キーワードとなりうる単語 (異なり数で 104 単語、総数で 4591 単語、但し複合語が多い) の認識率 (94.0%) が高くなっているためである。全体の音声認識率は検索性能にあまり影響しないということが言える。

4 キーワードの音声入力による検索実験

4.1 キーワードのグルーピング

前述の実験では、キーワードの入力がテキストであったが、音声での入力も考えられる。音声によるキーワード入力では、キーワードが認識された時、

1. キーワードが正しく認識された
2. キーワードが違う単語として認識された (異なる音節列)
3. 正しい音節列ではあるが，異なる語 (同音意義語) として認識された

という場合が考えられるが、機械には認識結果が正しいキーワードかどうかわからないので、どの場合も得られた認識結果を使って検索処理を開始せざるを得ない。同音意義語が存在する場合は、全ての同音異義語を使って検索する必要があり、同音意義語がない場合でも、認識尤度の高い認識結果候補単語を複数個使って検索する必要性も考えられる。いずれにしても、発声単語数よりも多い単語セット (キーワード候補) を使って検索処理が行なわれるため、必要以上の記事が検索されたり、また逆に全く記事が検索されない恐れがある。こういった不具合を解決する方法として、キーワード間の関連度を用いたキーワードの絞り込み手法を提案する。関連度とは、ある 2 つのキーワードがどれくらい関係しているかを表す尺度で、以下の値を用いる。

- 共起頻度の利用

  2 つの単語間の関連度を求める際に、ある記事において、ある単語とどの単語が同時に同じ記事に出現しやすいかという情報を用いる。

  2 つの単語をそれぞれ、W_1,W_2としたとき、これらの W_2の W_1に対する関連度 $R(W_1, W_2)$ を以下のように計算する。

$$R(W_1, W_2) = \frac{1}{2}\{\frac{f(W_1, W_2)}{f(W_1)} + \frac{f(W_1, W_2)}{f(W_2)}\}$$

  $f(W_i)$:W_iが出現した記事数 $(i = 1, 2)$

  $f(W_1, W_2)$:W_1,W_2が共に出現した記事数

- 相互情報量の利用

  相互情報量は、単語の共起や関連を客観的に表す尺度として用いられる。2 つの単語 W_1,W_2の相互情報量 $I(W_1; W_2)$ は、W_1と W_2を同じ記事で同時に観測する確率 $P(W_1, W_2)$ を、W_1と W_2を独立に観測する確率 $P(W_1)$,$P(W_2)$ と比較する。

$$I(W_1; W_2) = \log \frac{P(W_1, W_2)}{P(W_1)P(W_2)}$$

上記の式を変換して、

$$I(W_1;W_2) = \log \frac{\frac{f(W_1,W_2)}{N}}{\frac{f(W_1)}{N}\frac{f(W_2)}{N}}, \quad N：総記事数$$

2つの単語で、関連度が強いものはIの値が大きくなり、関連度がないものほど0に近づく。

ニュース記事から学習した前述の指標を使って、図2に示すように関連度の高いキーワード候補どうしをグルーピングする。関連度には閾値を設けており、この閾値を越える関連度をもつグループどうしを関連づけるわけである。この例は、6個のキーワードの候補がありうる場合を示している。矢印で結んであるキーワードどうしが関連度の高いキーワードで、1グループを形成している。ここでは3つのグループが作られているが、最もキーワードの数が多いG_1のグループを使って検索を行なう。

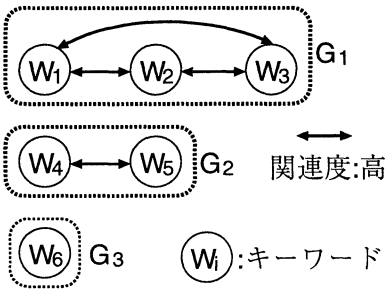

図2: キーワード候補のグルーピング

図3に実際の例を示す。これは、"公開"、"官庁"、"公務員"の3つのキーワードを入力したときの例である。"公開(こうかい)"の同音意義語として"更改"と"更改"、"官庁(かんちょう)"の同音意義語として"艦長"がある。この例では、"公開"—"官庁"間、"官庁"—"公務員"間、"公開"—"公務員"間の関連度が高くなっている(ある閾値を越えている)ので、これら3つのキーワード候補を1つのグループとしている。

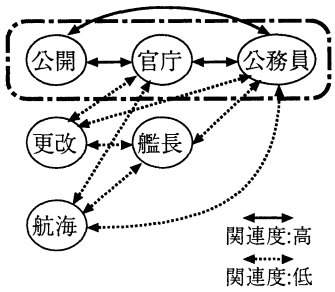

図3: グルーピングの実例

4.2 キーワードの音声認識

2名の話者にキーワードもしくはキーワード列を発話してもらい、認識実験を行なった。キーワード列とは、キーワードの連続のことであり、複合名詞などに該当する。発話してもらったキーワード(列)は、3節のテキスト入力の検索実験で用いたキーワード(列)と同じものである。

キーワード(列)の認識には、ニュース音声の時と同じ大語彙連続音声認識システム(語彙サイズは20000単語)を使用した。認識結果を表3と表4に示す。キーワード数は104個、キーワード列で数えると54個である。表3は、キーワードごとの認識率を求めた表で、表4は54個のキーワード列の内、どれだけが正しく認識できたかを示した表である。

表3: キーワードの音声認識率[%]

	置換率	挿入率	脱落率	正解率	正解精度
話者1	11.5	30.8	0.0	88.5	57.7
話者2	12.5	23.1	0.0	87.5	64.4

表4: キーワード列での認識結果
(注:括弧内数値は割合)
正解1：認識結果の1-bestのみの正解数
正解3：認識結果の3-bestまでの正解数
準正解：挿入を許す場合の正解数
(正解1, 正解3は、置換・挿入は不正解とする)

	正解1	正解3	準正解1	準正解3
話者1	29(53.7%)	32(59.3%)	45(83.3%)	48(88.9%)
話者2	26(48.1%)	29(53.7%)	42(77.8%)	44(81.5%)

表3、表4を見てもわかるが、比較的高い認識率が得られた。これは、キーワード列が複数の形態素から構成されており、認識時にbigramの言語確率がうまく機能しているためだと考えられる。これに対して、単一形態素のキーワードの認識は悪かった(69.6%)。

認識結果を調べて判明したこととして、キーワードの認識結果には余計な文字(とくに助詞)が挿入されてしまっているのがほとんどであるということ(表3の挿入率の高さからも言える)、正解のキーワード(列)はほぼ3-bestまでに入っている、ということが挙げられる。

キーワードの認識時、キーワードの脱落により検索結果が受ける影響に比べ、余計な単語(特に助詞)の挿入により受ける影響は少ないと思われる(余計な助詞、動詞などキーワードになり得ない単語(ストップワード)が挿入された場合は、キーワード候補を図1の検索部に入力する際に取り除くようにしているため)。つまり、表3の正解率と表4でいう準正解にあたる結果が、実質的なキーワード(列)の認識率に相当する。

表 5: Aデータベースに対する検索結果 (音声入力)

(a)1-best の結果のみを用いた場合

話者	一致率 [%]	湧きだし記事数	検索率 [%]
話者 1	83.1	30	76.7
話者 2	81.5	120	76.7

(b)3-best までの結果を用いた場合

関連度	話者	一致率 [%]	湧きだし記事数	検索率 [%]
共起頻度	話者 1	86.3	38	86.7
	話者 2	85.6	52	86.7
相互情報量	話者 1	86.3	32	90.0
	話者 2	85.6	39	86.7

表 6: Bデータベースに対する検索結果 (音声入力)

(a)1-best の結果のみを用いた場合

話者	一致率 [%]	湧きだし記事数	検索率 [%]
話者 1	75.8	80	73.3
話者 2	73.4	204	73.3

(b)3-best までの結果を用いた場合

関連度	話者	一致率 [%]	湧きだし記事数	検索率 [%]
共起頻度	話者 1	83.1	86	83.3
	話者 2	79.0	142	86.7
相互情報量	話者 1	83.9	73	90.0
	話者 2	79.8	60	86.7

4.3 検索実験

キーワードの音声認識結果を入力キーワードとして、検索実験を行なった。データベース (A) を使った場合と、データベース (B) を使った場合との検索実験で、検索結果にどれくらいの違いが現れるかを調べた。

キーワードの認識結果の 1-best のみを用いた場合と、3-best までを用いた場合とで検索実験を行なった。3-best の場合、同じ読みの単語や認識により発生した余計な単語も一緒にキーワード候補として入力した。例えば、"フィルム 史上" を入力したとすると、"史上 (しじょう)" と同じ読みの単語 ("市場"、"市上"、"試乗") を辞書から検索し、キーワード候補とする。このままでは、ユーザーが意図しない記事が検索されたり、また、欲しい記事が検索されないということになるので、4.1 節で述べたキーワード候補間の関連度 (共起情報、相互情報量) を用いてキーワード候補のグルーピングを行ない、そのグループ内の候補を使って検索を行なう。今回の検索実験では、複数のグループの中で一番多くの候補をもっているグループを使用した。これに対して、1-best の場合は認識結果そのものを入力とし、同音意義語は考えない。また、グルーピング処理も行なっていない。

ある記事に対するキーワード群を用いて (A),(B)2 種類のデータベースに対しての記事の検索結果を表 2, 表 5, 表 6に示す。表 2はデータベース (B) に対してキーワードがテキスト入力の場合、表 5はデータベース (A) に対してキーワードが音声入力の場合、表 6 はデータベース (B) に対してキーワードが音声入力の場合である。検索対象記事は全部で 30 記事で、一致率、検索率は全体の平均で、湧きだし記事数は全体の合計である。

表 5, 表 6で、キーワード (列) の認識率が高かったので、一致率、検索率とも比較的高い値になっている。しかし、余計な記事の湧きだしがかなり多くなった。また、表 2, 表 6のデータベース (B) に対しての検索結果に対しても同様なことが言えるが、両方とも音声の場合は、若干一致率が落ちている。これは、キーワードの認識が完全でないこと、グルーピングが必ずしもうまくいっていないことが考えられる。グルーピングは余計な単語を取り除くのには効果があるが、取り除きすぎ (関連度が低い場合) になる場合もある。また、正解でない候補どうしでグループを構成する場合もあった。

1-best のみ使用した場合と、3-best までを使用した場合とでは、一致率、検索率ともに 3-best までの方が 3〜10%程良くなっている。このことから、1-best のみを使うよりも 3-best までを使う方が良く、グルーピング手法が有効であると言える。

表 5, 表 6で、関連度の尺度を共起表現を用いた場合と、相互情報量を用いた場合での結果を載せているが、相互情報量を用いた方が、記事の湧きだしが少なくなっており、検索率も若干良い。一致率はほとんど変わらなかった。

5 おわりに

今回、ニュース音声データベースから、ニュース記事の検索システムを試作し、音声認識による書き起こしのデータベースを用いても検索能力が高いことを示した。また、キーワードが音声入力の場合の検索実験を行ない、音声入力でもグルーピングによりキーワード候補を絞り込むことで、高い検索率を得られることがわかった。しかし、多数の余計な記事が検索されているので、グルーピングの改良 (関連度の尺度など) が必要である。

本稿では、検索を行なう前にキーワード候補を絞り込んだが、その検索結果をさらに絞り込む方法として、音響的類似性などを使った方法を試みたい。

謝辞

この研究は、NHK 放送技術研究所のニュース音声データベース、ニューステキストデータベースを使わせていただいた。これらのデータベースを提供された NHK 放送技術研究所の関係諸氏に深く感謝する。

参考文献

[1] Dave Abberley,Steve Renals,Gary Cook: Retrieval of Broadcast News Documents with the THISL System, Proc.ICASSP,pp.3781–3784(1998.5)

[2] 福島, 赤峯:全文検索システム Retrieval Express の開発と評価, 言語処理学会, 第 3 回年次大会,pp.361–364(1997.3)

[3] 赤松, 花井, 甲斐, 峯松, 中川:新聞・ニュース文をタスクとした大語彙連続音声認識システムの評価, 情報処理学会, 第 57 回全国大会, pp.35–36(1998.10)