

RE-003

Webによるドキュメント拡張を利用した音声ドキュメント検索 Spoken Document Retrieval with Document Expansion Using Web

杉本 樹世貴[†]

Kiyotaka Sugimoto

西崎 博光[‡]

Hiromitsu Nishizaki

関口 芳廣[‡]

Yoshihiro Sekiguchi

1. はじめに

1990年代から、アメリカ国立標準技術研究所(NIST)とアメリカ国防総省が主催する情報検索関連の評価型ワークショップTREC(Text REtrieval Conference)が開催されている。1990年代後半から2000年までの間、TRECの研究分野(トラック)の1つとして、音声ドキュメント検索トラック(Spoken Document Retrieval Track)が開催され、世界中で音声ドキュメント検索の研究が行われてきた。現在は、ビデオ検索の評価ワークショップ(TRECVID⁴⁾)が開催されており、映像中に含まれる音声データを音声認識する等して、ビデオ検索精度の改善が図られている[1]。

一方、日本では、2006年に情報処理学会音声言語処理研究会のワークショップである「音声ドキュメント処理ワーキンググループ」が設立された。ワーキンググループは、『日本語話し言葉コーパス』(以下CSJと記す)を用いた音声ドキュメント検索用のクエリと、クエリに対する正解ドキュメントのセットを構築した[2]。

この評価セットはTREC SDR TrackやTRECVIDが採用しているニュース音声・映像とは異なり、話し言葉音声を対象としている。したがって、音声認識がより難しく、それに比例して検索精度も難しい。

この検索テストセットの構築により、日本でも大規模なデータを用いた情報検索技術開発の環境が整った。日本語の音声ドキュメント検索は、これまで、各研究機関が独自のデータベースを用いて研究を行ってきた[3, 4]。各研究機関が共通して利用できる日本語の評価セットができたことから、音声ドキュメント検索研究のさらなる発展が期待できる。

将来的に、音声やビデオといったマルチメディアデータが爆発的に増加していくことを考えると、これらの検索技術を開発することは喫緊の課題である。そこで本稿では、この検索テストセットに対する検索精度の改善を図ることを目的に、Webページを用いた新しい検索方法の提案を行う。

通常、音声ドキュメントを検索する場合、検索対象の音声データを音声認識することで音声波形を単語列など

にシンボル化し、そこからドキュメントに対するインデックスを作成し、そのインデックスを手掛かりにして、検索を行う。この際、音声認識処理を用いるため、音声認識誤りや未知語問題(音声認識辞書に必要な単語が登録されていない)が、検索精度に大きく影響する。

それらの問題を改善するため、検索対象のデータと内容が類似したWebページを音声ドキュメントのインデキシングに利用する。つまり、提案手法はWebデータを用いて音声ドキュメントの内容を拡張する。これによって、本来ならインデックスとして登録されて欲しい単語が、音声認識誤りや未知語が原因で、登録されない事態を回避するのが狙いである。

検索実験の結果、Webページを利用したドキュメント拡張により、検索精度を改善することができた。特に、未知語を含むクエリに対する検索精度が大きく改善している。このことから、提案手法の有効性を確かめられた。

2. 関連研究

音声ドキュメント検索問題を解決するために、90年代後半からこれまで、数多くの検索手法が開発されてきた。これは、前述したTRECの成果の賜物である。

例えば、Wechslerら[5]は、音素認識器を使って音声を書き起こし、入力検索語の音素表記とのマッチングを行っている。Ngら[6]は、音声の音素表記から作ったサブワードをインデキシングの単位として扱い、DPマッチングにより柔軟なマッチングを行っている。西崎ら[3]は、音声認識結果の誤りに対処するために、複数の音声認識システムを利用したり、音声認識システムに未知語検出処理を組み込むことで、音声認識システムで未知語と判定された区間のみを、サブワード単位のインデックスを構築する手法を提案した。岩田ら[4]は、語彙に影響されない音声ドキュメント検索として、サブワード単位の認識結果を利用している。また、クエリとドキュメントにおいて、高速にサブワード同士をマッチングさせる手法を提案している。

近年では、音声認識結果のラティス[7]やコンフュージョンネットワークを利用したり[8]、音声認識結果に信頼度を導入することで[9]、検索性能改善を図っている研究例もある。

日本語でもテストコレクションが整備されたことから、この検索テストコレクションを対象にした研究も行われている。まず、文献[2]では、テストコレクションを利用

[†]山梨大学大学院医学工学総合教育部, Educational Interdisciplinary Graduate School of Medicine and Engineering

[‡]山梨大学大学院医学工学総合研究部, Research Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

⁴⁾<http://www-nlpir.nist.gov/projects/trecvid/>

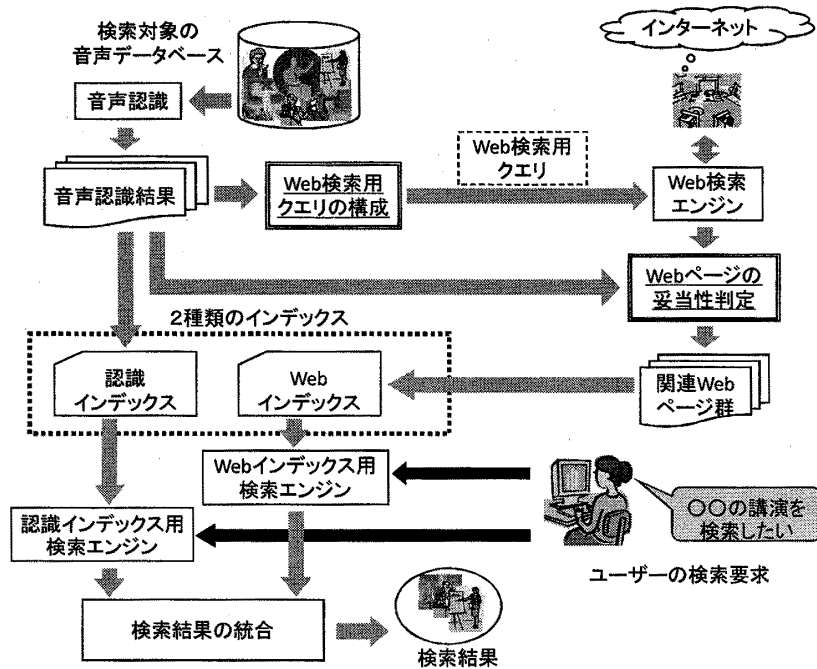


図1: Webを利用した音声ドキュメント検索処理の概要

した音声ドキュメント検索の基本的な検索精度が述べられている。また、秋葉らは [10], テキスト翻訳の技術を利用した語彙拡張を音声ドキュメント検索に応用し、テストコレクションでの有効性を示した。さらに、胡らは [8], 音声認識結果のコンフュージョンネットワークを利用することで、検索精度が改善できることを示した。前沢らは [11], 音声ドキュメントを認識する際に、様々な音声認識精度の書き起こしを用意し、認識精度が検索精度に与える影響や、サブワード (音節) をインデックスに用いることが、認識誤りや未知語に有効であることを示した。

これに対し、Webを利用したドキュメント拡張を行う利点として、2つの点を上げることができる。1点目は、音声認識誤りや未知語問題に対処できることである。2点目は、ユーザが欲する音声ドキュメント内に検索クエリーの単語が含まれていなくても、Webによるドキュメント拡張によりその単語が補完される点である。これにより、精度の高い検索が可能となる。

音声ドキュメント検索において、本研究と同様に Webを利用する研究も報告されている。例えば、寺尾ら [12] は、検索に用いる音声クエリと関連性のある Web の情報を利用することで、クエリ拡張を図っている。また宇野ら [13] は、音声ドキュメントと関連性のある Web 情報の件数と未知語率の関連性を研究している。

音声ドキュメント検索において、クエリ拡張を利用している研究として、Mamouら [14] が単語の発音に基づ

く拡張処理を提案している。

文献 [12] 等のように、クエリを拡張する手法に対して、本提案手法では、検索対象の音声ドキュメントの誤認識や未知語問題を解決する方法として、音声ドキュメントの拡張を Web ページによって行うことで検索性能の向上を目指している。

音声ドキュメント検索において、ドキュメント拡張を導入した研究としては、Singhalらの報告がある [15]。これは、ニュース音声を検索する際に、その音声と類似したデータをニュースコーパスから選択するという手法である。

今回、我々がターゲットにしているデータは、様々な話題を含んだ学会講演・模擬講演音声であり、これに類似したコーパスは今のところ存在しない。今後、様々な話題の音声 (映像) ドキュメントが増加していこうが、ドキュメント拡張の手法を導入する際に用いる類似コーパスとしては、いろいろな話題に対処しやすい Web が適していると考えられる。

そこで本稿では、Webを利用した音声ドキュメント拡張手法の有効性を確かめるため、検索実験を行ったので、それについて報告する。

3. Web ページを利用した音声ドキュメント検索処理

提案手法の処理の流れを図1に示す。

まず、検索対象の音声ドキュメントを大語彙連続音声

認識システムを用いて音声認識する。このとき、音声認識システムの認識辞書に登録されていない単語が未知語となる。クエリにこの未知語が含まれている場合、インデックスとマッチングができない。

この音声認識結果からストップワードを取り除き、インデックスを構築する。本稿では、これを“認識インデックス”と記す。

次に、Web ページから作成するインデックスについて説明する。まず、音声認識結果から、認識されたドキュメントと内容が類似した Web ページを検索するためのクエリを作成する。ここで構築する Web 検索用クエリの“質”によって、どれだけ内容が近い Web ページを収集できるかが決定される。したがって、クエリの生成方法にも工夫が必要である。しかし、本研究は Web ページによる音声ドキュメント拡張の有効性を示すことが目的であるため、Web 検索クエリの生成方法は単純な方法を採用した。これについては次節で述べる。

Web 検索クエリによって収集された Web ページの中には、音声ドキュメントの内容と無関係のものも多く含まれている。それらの Web ページは、認識結果との関連性の尺度に基づいて取り除いている。

このようにして集めた Web ページから、認識インデックスを作成した場合と同様にストップワードを取り除き、インデックスを構築する。本稿ではこれを“Web インデックス”と記す。

このように、音声ドキュメントに対して2種類のインデックスが構築される。つまり音声認識結果から直接作成される“認識インデックス”と、関連する Web ページから作成される“Web インデックス”である。

検索処理では、それぞれのインデックスに対してマッチング処理を行うため、2つの検索エンジンを用いる。ユーザの検索要求を2つの検索エンジンに入力し、それぞれのエンジンから得られた検索スコアを統合することで、最終的な結果を得る。

4. 関連する Web ページの取得

Web 検索を行うためには、音声ドキュメントからキーワード抽出を行い話題に合った適切な検索クエリを構成しなければならない。しかし、音声認識結果が形態素に分割されていること、話題の特定に不向きな単語（機能語や一般語）が多く含まれていることから、話題に合った検索クエリを構成することは簡単ではない。

そこで、完璧ではないが、ある程度適切な Web ページを収集できるように、Web 検索用クエリの構成方法を検討した。

4.1 Web 検索用クエリの構成と検索

Web 検索用クエリを構成するために、音声ドキュメントの音声認識結果中に含まれる単語の品詞を判定し、適

例文

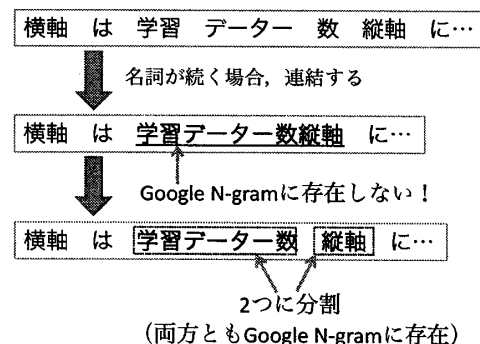


図 2: 名詞 N-gram 作成の例

宜単語の連結と削除を行うことで、実在する名詞 N-gram を抽出する。この名詞 N-gram を Web 検索用クエリとする。

具体的には、以下の手順で行う。

1. 名詞 N-gram の作成 : 話題を特定する単語 (名詞) が続く限りそれらを連結し、名詞 N-gram を作成する。N の大きさは任意である。このとき、Web 日本語 N-gram 第一版 (“Google N-gram” と記す)⁵⁾ を使用し、連結した名詞 N-gram が実在するかを確認する。連結された名詞 N-gram が Google N-gram に存在すれば、Web 検索用クエリの候補とする。もし、存在しなければ、連結した名詞 N-gram を解体し、再度 Google N-gram に存在するかどうかを確認する。この処理を繰り返すことで、音声認識結果のテキストから名詞 N-gram を構築する。具体的な処理の流れの例を図 2 に示す。
2. クエリの構成 : 作成した名詞 N-gram の出現頻度をカウントし、頻度が高い順に 5 つの名詞 N-gram を採用する。この 5 つの名詞 N-gram すべてを Web 検索用クエリとして、Web 検索エンジンに入力する。

以上の手順で作成した Web 検索用クエリを入力とし、“Yahoo! Web 検索 API”⁶⁾ を使って音声ドキュメントに関連する Web ページ群を収集する。収集する Web ページ数は、1 音声ドキュメントに対して 50 ページとした。Web 検索クエリによっては、50 ページの収集が困難な場合がある。その場合は、最も頻度の低い名詞 N-gram をクエリから取り除き、再度検索を行う。

4.2 Web ページの妥当性判定

収集された Web ページの中には、音声ドキュメントの内容と無関係のページも多く含まれている。

⁵⁾ <http://www.gsk.or.jp/catalog/GSK2007-C/catalog.html>

⁶⁾ <http://developer.yahoo.co.jp/webapi/search/>

そこで、収集した Web ページと音声ドキュメントの認識結果との間の類似度を尺度として用いることにより、インデックスとして使用する Web ページの取捨選択を行う。

この類似度計算にはコサイン尺度を用いる。これは、各ドキュメントを、そのドキュメントに含まれる単語の出現頻度を要素としたドキュメントベクトルで表現し、ドキュメントベクトル間の成す角を求めることにより類似度を求めるものである。コサイン尺度の計算式を式(1)に示す。

$$\cos(\mathbf{WD}_i, \mathbf{R}) = \frac{\sum_{j=1}^m WD_{ij} \cdot R_j}{\sqrt{\sum_{j=1}^m WD_{ij}^2} \sqrt{\sum_{j=1}^m R_j^2}} \quad (1)$$

ここで、 \mathbf{R} は音声ドキュメントの認識結果、 R_j は認識結果中の単語 w_j の出現頻度、 WD_{ij} は Web ドキュメント \mathbf{WD}_i 中の単語 w_j の出現頻度、 m は \mathbf{R} と \mathbf{WD}_i に含まれる単語の種類数を表す。

妥当性評価を行った結果、インデックスとして使用する Web ページが 0 件となった音声ドキュメントに対しては、認識インデックスで代用する。

5. 検索システム

5.1 検索手法

音声ドキュメントの検索エンジンには、汎用連想計算エンジン GETA⁷⁾[16] を利用した。GETA を利用することで、大規模なドキュメント-単語集合間の類似度を高速に計算することが可能である。

インデックスは形態素単位とする。先行研究[17]では、形態素の他に文字バイグラム、および形態素と文字バイグラムの組み合わせが試されており、その効果が示されている。

認識インデックスおよび Web インデックスを構築する際に、ストップワードを除去して内容語のみをインデキシングする。本研究では、内容語として、名詞、動詞、形容詞、形容動詞を用い、これら以外の品詞をストップワードとする。

音声ドキュメントと検索クエリ間の類似度計算には、GETA に用意されている SMART 法⁸⁾[18] を採用した。インデックスは、ドキュメント長で正規化された TF・IDF 法により重み付けされている。ベクトルで表現されたドキュメントとクエリから、余弦計算に基づいてドキュメントとクエリ間の類似度を計算する。

5.2 検索結果の統合

認識インデックスと Web インデックスのそれぞれのインデックスを用いたときの検索結果を統合し、最終的

な結果を得る。

2つのインデックスを用いて検索された音声ドキュメント D の最終的なスコア $sim(D)$ は、式(2)に示すように、認識インデックスの検索結果の検索スコアと Web インデックスの検索結果の検索スコアの線形補間により計算される。

$$sim(D) = (1 - \alpha) \times sim(D|R) + \alpha \times sim(D|W) \quad (2)$$

ただし、 $sim(D|R)$ は認識インデックス、 $sim(D|W)$ は Web インデックスを用いたときの音声ドキュメント D の検索スコアとなる。 α は、各インデックスに対する重み係数である。この α を、0.0 から 1.0 までの 0.1 刻みで変化させ検索実験を行う。

また、検索クエリに対する未知語率を OOV と定義したとき、この OOV を重み係数として用いることで(つまり、 $\alpha = OOV$)、クエリ毎に動的に重みを変化させる手法も実験する。ここで用いる未知語率 (OOV) とは、音声ドキュメントを検索するためのクエリに含まれる単語のうち、音声認識辞書に含まれていない単語の割合である⁹⁾。

6. 検索実験

6.1 テストコレクション検索条件

検索対象のデータには、CSJ を対象に構築された検索評価用テストコレクションを用いる。

6.1.1 テストコレクションについて

テストコレクションには、CSJ に収録されている「学会講演」「模擬講演」の 2702 講演(約 600 時間以上の音声データ)に対し、それら講演、または講演の一部を検索するための検索クエリと正解の講演データのセットが収録されている。

正解リストには、クエリに対して適合している講演 (Relevant)、部分的に適合している講演 (Partially Relevant) が含まれている。クエリ数は全部で 39 個である。

6.1.2 音声認識

検索テストコレクションには、2702 の講演音声データが存在する。これらを音声認識することで、認識インデックスと Web インデックスを構築しなければならない。以下、講演音声を認識する条件について説明する。

現状では、2702 講演を音声認識するために必要なモデルを完全にオープンな学習データから学習することは難しい¹⁰⁾。例えば、文献[2]でも、CSJ の講演音声を認

⁷⁾ <http://geta.ex.nii.ac.jp/>

⁸⁾ <http://geta.ex.nii.ac.jp/geta.N2002/gdoc/geta/tutorial/libae/section5.html>

⁹⁾ 音声ドキュメントに対する未知語率でないことに注意されたい。

¹⁰⁾ ただし、クロスバリデーションを行えば、膨大な時間がかかるがオープンなモデルを構築することは可能

表 1: 言語モデルの語彙サイズ別の音声認識率と未知語率

語彙サイズ	単語正解率 [%]	正解精度 [%]	クエリに対する未知語率 [%]	未知語を含むクエリ数
7k	74.7	70.1	16.2	14
12k	75.8	70.1	14.8	13
17k	76.9	71.6	11.8	11

識する際に用いる音響モデル, 言語モデルの学習には, クローズなデータを含んでいる. したがって, 80%以上の高い音声認識精度が得られている.

我々の実験では, 音声認識に用いる音響モデル, 言語モデルの学習データができるだけオープンになるように, 文献 [2] で用いられている各モデルの学習データよりも少なくしている.

本研究において, 音声認識に用いた音響モデルは, CSJ 学会講演から認識評価用セットを除いた学会講演 970 講演から学習したトライフォンである. 使用する特徴量は, 12次元の MFCC, Δ MFCC, $\Delta\Delta$ MFCC と, Δ パワー, $\Delta\Delta$ パワーの全 38 次元となっている.

言語モデルは, 音響モデルの学習データと同じ学会講演 970 講演から, トライグラムを学習した. 本研究では, 意図的に 3 種類の語彙サイズ (7k, 12k, 17k) の言語モデルを用意した. これは, 語彙サイズを変えることで, 音声ドキュメントの音声認識率や未知語率を変化させ, 様々な認識率での提案手法の有効性を示すためである.

音声認識のデコーダには Julius ver.4.1 を用いている.

CSJ 2702 講演に対する音声認識を行った結果を, 表 1 に示す. 表 1 において, クエリに対する未知語率とは, 5.2 で述べたものと同じで, 音声ドキュメントを検索するためのクエリに含まれる単語のうち, 音声認識辞書に含まれていない単語の割合である. 未知語を含むクエリ数とは, 全 39 個のクエリに対し, 未知語を含んでいるクエリの個数である.

表 1 を見ると, 当然ながら語彙サイズが小さい言語モデルを使うと, 音声認識の精度は悪化している. また, クエリに対する未知語率や未知語を含むクエリ数は増加している.

6.1.3 検索単位

本研究では, 検索の単位は講演単位とする. すなわち, 1 ドキュメント 1 講演である.

検索されたドキュメントの適合性の判定は, "適合" はもちろんのこと, "部分適合" しているドキュメントも正解とする.

6.2 評価尺度

検索精度を計る尺度には, 補間 11 点平均精度 (Interpolated 11-points Average Precision, "11-pt AP" と記

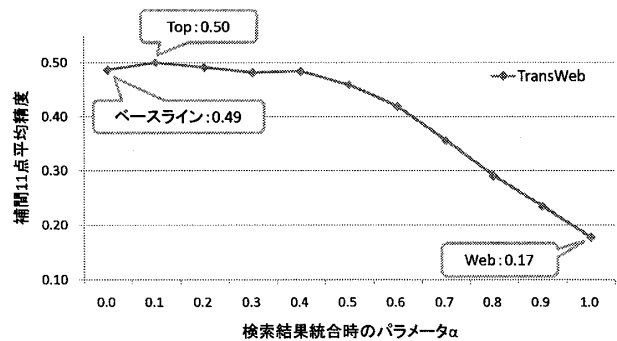


図 3: 線形結合の重み α を変化させたときの補間 11 点平均精度の変化 (書き起こしデータ, 認識率 100%)

す) を用いる.

これは, 再現率 (Recall) が 0.0, 0.1, 0.2, ..., 1.0 のときの補間精度 (Precision) を平均したものである. 再現率は 0.0~1.0 までの 11 点であるので, 11 点の補間精度を平均する. 11-pt AP の計算式を式 (3) に示す [17].

$$IP(x) = \max_{x \geq R_i} P_i, \quad AP = \frac{1}{11} \sum_{i=0}^{10} IP\left(\frac{i}{10}\right) \quad (3)$$

ここで, R_i と P_i は, それぞれ, 検索結果の上位 i 番目までに着目したときの再現率と精度である.

検索実験では, 上位 1000 件までの結果を出力し, 上記の式により評価する.

6.3 実験結果

Web ページを用いたドキュメント拡張の効果を, 検索実験により評価した. 結果を図 3, 図 4, 図 5 に示す.

6.3.1 書き起こしドキュメントに対する結果

図 3 は, 音声ドキュメントを音声認識システムではなく, 人手により書き起こしたドキュメント集合を対象にした検索実験結果である. すなわち, 音声認識誤りを含んでいないため (認識率 100%), 純粋なテキスト検索結果となる. 検索対象のドキュメントに対する類似 Web ページは 4. 節で述べた方法で集められている.

図 3 を見ると, $\alpha = 0.0$ のとき, すなわち Web によるドキュメント拡張を行っていない結果 (ベースライン) に比べ, ドキュメント拡張を行った方が検索結果が良くなっ

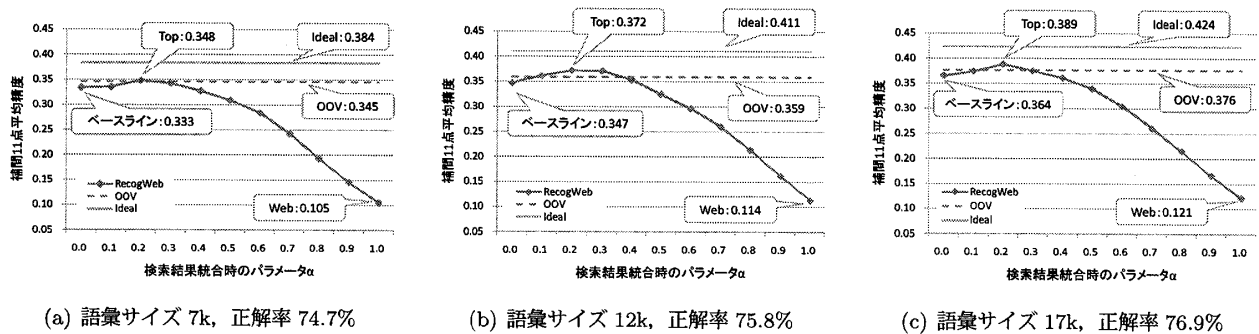


図 4: 全 39 クエリに対する線形結合の重み α を変化させたときの補間 11 点平均精度の変化 (音声認識率別)

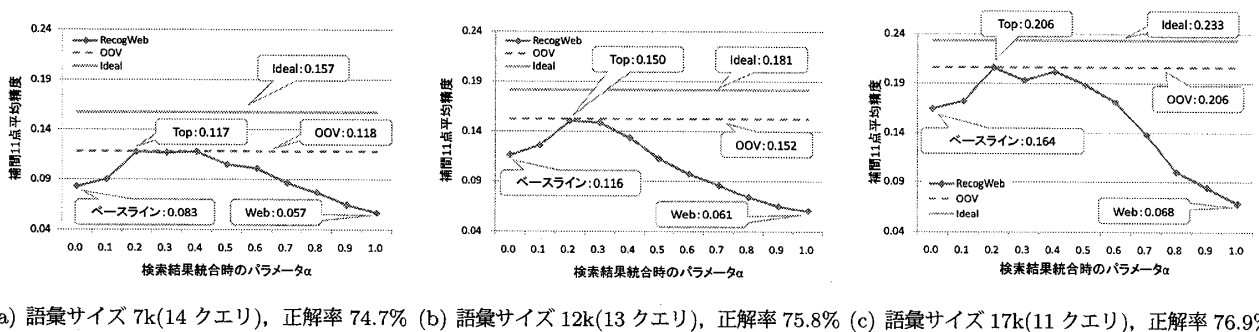


図 5: 未知語を含むクエリのみに対する線形結合の重み α を変化させたときの補間 11 点平均精度の変化 (音声認識率別)

ている。 $\alpha = 0.1$ のときに、ベースラインよりも 0.013 ポイントの改善が得られている。この結果より、音声認識誤りのないテキスト検索においても、Web によるドキュメント拡張の効果があることが示された。

6.3.2 音声認識率別の結果

続いて、図 4(a)~(c) は、全 39 個の検索クエリに対する音声ドキュメントの認識率別の検索結果を示している。

“RecogWeb” は、音声ドキュメントを Web ページで拡張し、統合時に重みを変化させたときの結果を表している。図 3 と同じで、 $\alpha = 0.0$ のときは、Web によるドキュメント拡張を行っていない結果で、これをベースラインとする。一方で、 $\alpha = 1.0$ のときは認識インデックスを使わず、ドキュメント拡張に用いた Web ページだけを検索に用いた結果である (“Web” と表記)。また、図 4 では、すべてのクエリに対して共通の重み α を設定したときに、最も検索精度が良い結果 (“Top”), 各クエリ毎に最適な重みを設定したときの結果 (最適な統合が行われたときの理想値: “Ideal”) も示している。なお、語彙サイズ 7k の結果のみ、類似度による Web の適合性判断を行っている。

図 4 の結果を見ると、言語モデルの語彙サイズの低下

に伴う音声認識率の低下により、検索精度が全体的に悪化している。しかし、すべての認識インデックスにおいて、Web によるドキュメント拡張の効果が見られ、ベースラインに比べて検索精度を改善することができた。どのインデックスにおいても、 $\alpha = 0.2$ のときに最適な結果となった。

また、図中の “OOV” は、式 (2) において、 α の値をクエリ毎の未知語率を用いて動的に決定したときの結果である。(a) の言語モデルの語彙サイズが 7k のときは、各クエリに対する未知語率を用いた動的な重み設定の実験結果と、一律に重みを決定したときの結果がほぼ同じとなった。しかし、12k と 17k のときは、一律の重み設定の結果よりも、若干劣る。理想値とも相当の開きがあることから、未知語率による動的な重み設定は一定の改善はみられるものの、改善の余地がある。

次に、図 4 が全てのクエリに対する結果だったのに対し、図 5 は、音声認識辞書に対する未知語を含んでいるクエリに対する実験結果となっている。言語モデルの語彙サイズが異なると含まれる未知語率も変化するため、サイズによってクエリ数が異なっている (括弧内数値を参照)。

この結果を見ると、どの認識インデックスで比べても、

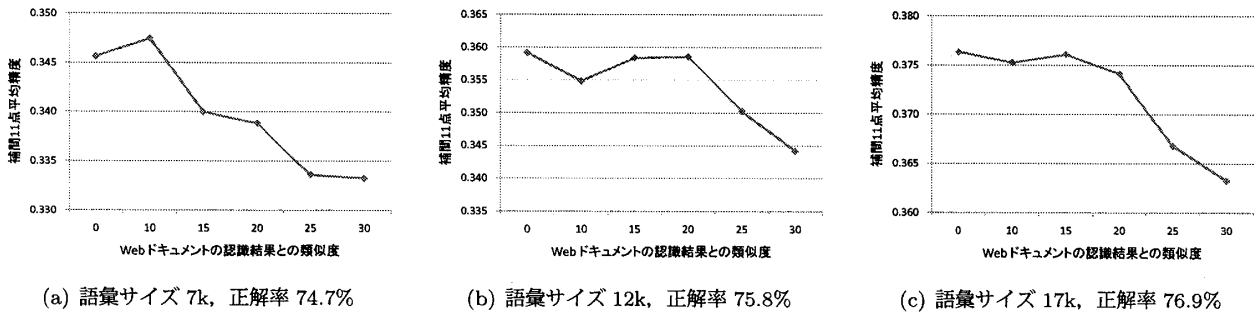


図 6: Web ページの妥当性判定を取り入れたときの検索性能 (全 39 クエリ)

Web ドキュメント拡張を行うことでベースラインよりも大幅に検索精度が改善していることが分かる。改善幅は、図 4 の 39 クエリに対する結果よりも大きいことから、本提案手法は、特に、未知語を含む検索クエリに対して有効であると言える。すなわち、Web ドキュメント拡張により未知語を補完できている。また、図 5 では、どの認識率においても、未知語率による動的な重み決定が、一律の重み設定よりも優れている。未知語を含むクエリに対しては、未知語率による重み決定が有効であった。

以上の結果から、音声ドキュメント検索において、Web ページによるドキュメント拡張の有効性を示すことができた。特に、図 5(c) においては、ドキュメント拡張により検索精度は 0.042 ポイントの改善が見られた。ただし、最適な重みを選択したときの理想値の精度と比べてみると、まだ改善の余地が残っているため、今後は動的な重み決定手法を工夫する必要がある。

6.3.3 Web ページの妥当性判定

図 6 は、Web ページの妥当性判定を行った場合の結果、すなわち Web インデックスとして使用する Web ページを、認識結果との類似度に基づいて選択したときの検索精度を示している。

図 6 を見ると、言語モデルの語彙サイズが 7k では、類似度の閾値を 10% としたときすなわち、類似度が 10% 以上の Web ページを使用した場合、に最も良い検索結果が得られた。しかし、語彙サイズ 12k、17k においては、類似度の閾値が 0% のとき、つまり収集した Web ページ全てをインデックスとして用いた場合、検索精度は最も良かった。

Web ページに対するフィルタリング効果はあまり見られなかったが、これは Web 検索用クエリを比較的簡単な手法により構成したことにより、音声ドキュメントの内容と無関係の Web ページが多く集まってしまったこと、また、類似度計算の方法も単語の出現頻度のみを使用していること等が原因であると考えられる。今後は、

より適した Web ドキュメントを収集する方法を開発することで、Web ページの妥当性判定処理が効いてくるのではないかと考えている。

6.3.4 音節インデックスとの比較

本研究では、単語単位のインデックスを利用してきた。先行研究 [11] では、同じ検索テストセットを利用し、文字や音節等の異なるインデックス単位を利用し、認識誤りや未知語に頑健な手法を提案している。この文献 [11] の手法と比較する。

まず、単語インデックスに加え、文字 bigram をインデックスに追加することで、11 点平均精度で 0.035 ポイントの改善を図ることができている。さらに、単語と文字 bigram インデックスと音節インデックスを併用することで、単語のみのインデックスに比べて 0.041 ポイント改善している。

本稿では、Web によるドキュメント拡張を行うことで、単語インデックスに比べて 0.025 ポイントの改善を図ることができた。しかし、認識インデックスと Web インデックスの補間重みを最適に選択すると、0.060 ポイントの改善が行えることから、Web 拡張の利用には改善の余地がある。

また、6.3.3 節で述べたように、今回の実験において自動で収集した Web ページに不適切なものも多かったこともあり、Web ページ収集方法を改善することで、さらに改善が見込める。また、先行研究 [11] で用いた音節インデックスや文字 bigram を併用することで、さらなる改善が期待できそうである。

7. おわりに

本稿では、音声ドキュメント検索タスクにおいて、検索対象の音声ドキュメントと類似性の高い Web ページを用いてドキュメント拡張を行う有効性について論じた。

音声ドキュメント検索用テストコレクションを用いて評価実験を行った結果、音声認識結果のみをインデック

スとした場合よりも、Web ページを用いることで検索精度を改善できることを示した。提案手法では、音声ドキュメントの音声認識率が異なるインデックスを3種類用意し検索実験を行ったが、いずれの認識率の場合においても、Web による検索精度の改善が見られた。

テストコレクションの全39個の検索クエリに対する実験結果では、補間11点平均精度において最大で2.5%の改善、さらに、未知語を含む検索クエリに対しては、最大で4.2%の検索精度の改善を確認できた。

今後は、Web 検索用クエリの“質”がWeb インデックスの“質”を決める上で重要となってくるため、Web 検索用クエリの構成法の再検討を行い、Web インデックスの“質”を改善していく。また、2つのインデックスを統合する方法には、まだ十分な改善の余地が残されているため、統合時の動的な重みの決定方法などを検討し、検索精度を改善していく予定である。

参考文献

- [1] Y.J. Cheng and H.H. Chen, “Aligning words from speech recognition and shots for video information retrieval,” Proc. of the TRECVID2004, 2004.
- [2] T. Akiba, K. Aikawa, Y. Itoh, T. Kawahara, H. Nanjo, H. Nishizaki, N. Yasuda, Y. Yamanashita, and K. Itou, “Construction of a test collection for spoken document retrieval from lecture audio data,” IPSJ Journal, vol.50, no.2, pp.1234-1245, 2009.
- [3] 西崎博光 中川聖一, “音声認識誤りと未知語に頑健な音声文書検索手法,” 電子情報通信学会論文誌, vol.J86-D-II, no.10, pp.1369-1381, 2003.
- [4] 岩田耕平, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “語彙フリー音声文書検索手法における新しいサブワードモデルとサブワード音響距離の有効性の検証,” 情報処理学会論文誌, vol.48, no.5, pp.1990-2000, 2007.
- [5] M. Wechsler, E. Munteaun, and P. Schauble, “New techniques for open-vocabulary spoken document retrieval,” Proc. of ACM SIGIR’98, pp.20-27, 1998.
- [6] K. Ng and V.W. Zue, “Subword-based approaches for spoken document retrieval,” Speech Communication, vol.32, no.3, pp.157-186, 2000.
- [7] Y. cheng Pan, H. lin Chang, B. Chen, and L. shan Lee, “Subword-based position specific posterior lattices (S-PSPL) for indexing speech information,” Proc. of the INTERSPEECH 2007, pp.318-321, 2007.
- [8] 胡新輝, 吳友政, 柏岡秀紀, “Confusion network を用いた音声ドキュメントの検索及び評価に関する研究,” 第2回音声ドキュメント処理ワークショップ講演論文集, pp.85-90, 豊橋技術科学大学メディア科学リサーチセンター, 2008.
- [9] W. Kim and J.H.L. Hansen, “Advances in speechfind: Transcript reliability estimation employing confidence measure based on discriminative sub-word model for sdr,” Proc. of the INTERSPEECH 2007, pp.2409-2412, 2007.
- [10] T. Akiba and Y. Yokota, “Spoken document retrieval by translating recognition candidates into correct transcriptions,” Proc. of the INTERSPEECH 2008, pp.2166-2169, 2008.
- [11] 前沢慎吾, 杉本樹世貴, 西崎博光, 関口芳廣, “音声ドキュメント検索用テストコレクションにおける音節インデックスの効果,” 第71回全国大会講演論文集第2分冊, E3-10, 情報処理学会, 2009.
- [12] M. Terao, T. Koshinaka, S. Ando, R. Isotani, and A. Okumura, “Open-vocabulary spoken-document retrieval based on query expansion using related web documents,” Proc. of the INTERSPEECH 2008, pp.2171-2174, 2008.
- [13] 宇野有, 伊藤仁, 伊藤彰則, 牧野正三, “音声ドキュメントの索引付けに向けたウェブ検索を用いたデータ収集における未知語率の検討,” 春季講演論文集, 3-Q-30, pp.275-276, 日本音響学会, 2009.
- [14] J. Mamou and B. Ramabhadran, “Phonetic query expansion for spoken document retrieval,” Proc. of the INTERSPEECH 2008, pp.2106-2109, 2008.
- [15] A. Singhal and F. Pereira, “Document expansion for speech retrieval,” Proc. of ACM SIGIR’99, pp.34-41, 1999.
- [16] 高野明彦, 西岡真吾, 丹羽芳樹, “連想に基づく情報アクセス技術: 汎用連想計算エンジン geta を用いて,” 情報の科学と技術, vol.54, no.12, pp.634-639, 12 2004.
- [17] 秋葉友良, 相川清明, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田宜仁, 山下洋一, 伊藤克亘, “音声ドキュメント検索テストコレクションの試作と基本検索性能評価,” 第1回音声ドキュメント処理ワークショップ講演論文集, pp.73-80, 豊橋技術科学大学メディア科学リサーチセンター, 2007.
- [18] A. Singhal, C. Buckley, and M. Mitra, “Pivoted document length normalization,” Proc. of ACM SIGIR’96, pp.21-29, 1996.