

E-013

## 音声ドキュメント検索における分割したドキュメントを用いる Web 収集の効果

## Effect of Collecting Web Using Segmented Documents on Spoken Document Retrieval

杉本 樹世貴<sup>†</sup>西崎 博光<sup>‡</sup>関口 芳廣<sup>‡</sup>

Kiyotaka Sugimoto

Hiromitsu Nishizaki

Yoshihiro Sekiguchi

## 1. はじめに

音声ドキュメント検索を効率良く行うには、音声認識結果をテキスト検索と同様に処理することが望ましい。しかしその場合、音声認識誤りや未知語が問題となり検索精度の低下を引き起こしてしまう。

我々はこれまでに、Web によるドキュメント拡張を利用することで、それらの問題に対処し検索精度を改善してきた [1]。

今回は、更なる検索精度の改善を目標として、1 講演内にも複数の話題が出現することに着目して、分割したドキュメントを用いた Web 収集方法を提案する。

検索実験の結果、分割したドキュメントを用いて Web 収集を行うことで検索精度を改善することができたので、提案手法について報告する。

## 2. セグメント分割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 (“Web 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を複数のセグメントに分割する。そして、各セグ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構成し、Web ページを収集する。

このようにして収集した 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、各セグメント毎にインデックスを構築する。すなわち 1 音声ドキュメントに対し、セグメント分割数分の Web インデックスが存在する。

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。ユーザの検索クエリを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。

## 3. 検索システム

## 3.1 Web 検索用クエリの構成と検索

Web 検索用クエリには、セグメント毎に名詞 N-gram の作成 [1] を行い、式 (1) により求めたスコアの上位 5 種類の名詞 N-gram を採用する。

$$Score(w_i) = \frac{1 + tf_i \times \log_{10}(1 + N)}{\sum_{k=1}^m tf_k} \quad (1)$$

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

<sup>‡</sup>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

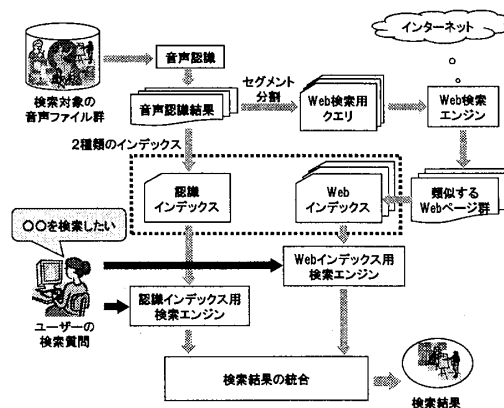


図 1: セグメント分割を利用した音声ドキュメント検索処理の概要

$tf_i$  は N-gram  $w_i$  の出現頻度、 $N$  は N-gram  $w_i$  を構成している単語数、 $m$  は、セグメント内の N-gram 種類数である。

Web 検索には、作成した Web 検索用クエリを入力とし、“Yahoo! Web 検索 API”<sup>1)</sup> を使ってセグメント毎に関連する 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。すなわち、分割数が 5 のとき、1 セグメントに対して 10 ページとなる。

## 3.2 検索手法

検索システムには、汎用連想計算エンジン GETA[2] を利用した。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 法を採用した。インデックスは、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。

## 3.3 検索結果の統合

認識インデックスと Web インデックスにより得られた検索結果を統合し、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。

$$sim(d) = (1 - \alpha) \times sim(d|r) + \alpha \times \max_{s \in S} sim(d_s|w) \quad (2)$$

ここで、ドキュメント  $d$  の  $n$  分割したセグメント集合を  $S = \{d_1, d_2, \dots, d_n\}$  とする。 $sim(d|r)$  は認識インデックスでの音声ドキュメント  $d$  の検索スコア、 $sim(d_s|w)$  は Web インデックスでの音声ドキュメント  $d$  に対するセグメント  $s$  の検索スコアとなる。 $\alpha$  は、各インデックスに対する重み係数である。この  $\alpha$  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。

<sup>1)</sup> <http://developer.yahoo.co.jp/webapi/search/>

表2: 各セグメント分割数に対する全39クエリでの補間11点平均精度

セグメント 分割数 (n)	$\alpha$												Ideal
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	OOV	
1	0.364	0.373	0.378	0.374	0.374	0.365	0.338	0.297	0.240	0.194	0.157	0.383	0.430
3	0.364	0.373	0.381	0.377	0.373	0.360	0.335	0.311	0.269	0.222	0.163	0.379	0.427
5	0.364	0.366	0.375	0.375	0.378	0.373	0.356	0.320	0.273	0.223	0.171	0.372	0.441
7	0.364	0.370	0.371	0.374	0.371	0.358	0.343	0.321	0.279	0.217	0.155	0.384	0.432
10	0.364	0.370	0.378	0.378	0.386	0.378	0.362	0.331	0.278	0.214	0.154	0.383	0.446

表3: 各セグメント分割数に対する未知語を含む11クエリのみでの補間11点平均精度

セグメント 分割数 (n)	$\alpha$												Ideal
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	OOV	
1	0.164	0.189	0.203	0.227	0.245	0.243	0.235	0.206	0.179	0.161	0.138	0.229	0.276
3	0.164	0.178	0.207	0.199	0.202	0.210	0.181	0.155	0.145	0.121	0.097	0.217	0.242
5	0.164	0.162	0.169	0.172	0.197	0.206	0.197	0.170	0.160	0.147	0.125	0.191	0.255
7	0.164	0.176	0.191	0.202	0.224	0.213	0.216	0.206	0.182	0.150	0.128	0.233	0.273
10	0.164	0.175	0.207	0.202	0.221	0.227	0.224	0.221	0.195	0.172	0.145	0.228	0.288

表1: 検索テストコレクションの音声認識率と未知語率

Corr. [%]	Acc. [%]	OOV [%]	Query
76.9	71.6	11.8	11

#### 4. 検索実験

分割したセグメント毎に Web 検索用クエリを構成することの有効性を検索実験により評価した。

##### 4.1 実験条件

実験条件を以下に示す。

**テストコレクション** テストコレクションには、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た。

CSJに収録されている「学会講演」「模擬講演」の2,702講演に対し、それらの講演、または講演の一部を検索するための検索クエリ39個と正解の講演データのセットが収録されている[3]。

検索の単位は、講演単位とする。すなわち検索されたセグメントを含む講演が1ドキュメントとなる。

**音声認識** 検索テストセットに対して行った音声認識結果を表1に示す。音声認識のデコーダには Julius rev.4.1を用いている。音響モデルは、特徴ベクトルが38次元のtriphone、言語モデルは、語彙サイズが17kのtrigramを使用する。

表1において、“OOV”とは、検索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。“Query”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。

**セグメント分割数** 今回は、分割したドキュメントを用いて Web 収集を行うことの有効性を示すことが目的であるため、簡単な方法として、発話数を基にドキュメントを  $n$  等分 ( $n = 3, 5, 7, 10$ ) する。

**評価尺度** 検索精度を計る尺度には、検索結果の上位1,000ドキュメントに対する、補間11点平均精度を使用する。

##### 4.2 実験結果

実験結果を表2、表3に示す。表2は、全39個の検索クエリ、表3は、未知語を含む検索クエリに対するセグメント分割数別の検索結果である。

$n = 1$  は、1講演単位で作成された Web インデックスを用いたときの結果で、本研究のベースラインである。また、“OOV”は、各検索要求に対する未知語率を用いて動的に  $\alpha$  を設定し

たときの結果であり、“Ideal”は、各検索要求毎に最適な重み  $\alpha$  を設定したときの結果である。

表2を見ると、すべてのクエリに共通の重みを設定したときに検索精度が最も良いのは、セグメント分割を行った場合の0.386 ( $n = 10, \alpha = 0.4$ )であり、ベースラインの0.378 ( $n = 1, \alpha = 0.2$ )より高い値となっている。また、“OOV”や“Ideal”についてもセグメント分割を行うことで、検索精度が改善されている。

次に、表3を見ると、すべてのクエリに共通の重みを設定したときや“OOV”について、セグメント分割を行うことで検索精度が悪化している。しかし、“Ideal”を見ると、セグメント分割を行った場合 ( $n = 10$ ) が最も高い値をとっている。これは、セグメント分割を行った方が理想的な統合が行われた場合の改善率が高いことを示している。

このことから、分割したセグメント毎に Web ページを収集することの有効性が示された。ただし、“Ideal”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。

#### 5. おわりに

本稿では、音声ドキュメント検索テストセットを用いた評価実験を行うことで、分割したセグメント毎に Web 検索用クエリを構成することの有効性について論じた。

評価実験の結果、1講演単位で Web 収集を行う場合よりも、分割したドキュメントを用いて Web 収集を行うことで、検索精度を改善できることを示した。

検索テストコレクションの全39クエリに対する実験結果では、理想的な統合を行うことで、セグメント分割を行わない場合に対し、1.6%の改善を確認できた。さらに、未知語を含むクエリに対しては、1.2%の改善を確認できた。

今後は、より話題にあったセグメント分割を行うために、動的セグメント分割手法の検討、理想的な統合を行うための重みの決定方法について検討を行い、検索精度の改善をしていく予定である。

#### 参考文献

- [1] 杉本樹世貴 他, “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,” 第8回情報科学技術フォーラム講演論文集, RE-003, pp.27-34, 2009.
- [2] 高野明彦 他, “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, 2004.
- [3] T. Akiba et al, “Construction of a test collection for spoken document retrieval from lecture audio data,” IPSJ Journal, vol.50, no.2, pp.1234-1245, 2009.