

広域文書類似度と局所文書類似度を用いた 講演音声ドキュメント検索

南條 浩輝^{1,a)} 弥永 裕介¹ 吉見 毅彦¹

受付日 2011年9月26日, 採録日 2012年3月2日

概要: 講演音声集合から探したい内容を表す 30 秒から 1 分程度の箇所を検索する方法について述べる. 本論文では, 講演の各部分には講演全体が扱うトピックのサブトピックがある可能性が高いことに着目し, 講演単位などの大きなまとまり (広域文書) での粗い検索と, 30 秒から 1 分程度の単位 (局所文書) の検出したい単位での詳細な検索とを階層的に統合する方法を提案する. 具体的には, トピック推定は行わずに発話数に基づいて分割した局所文書と広域文書からなる階層構造を自動生成したうえで, クエリと広域文書との類似度を局所文書との類似度に統合して, 局所文書を検索する方法を提案する. CSJ 音声ドキュメントコレクションの 15 発話単位, 30 発話単位および 60 発話単位の検索タスクにおいて, 検索精度 (11 点平均精度) の有意な向上が確認できた.

キーワード: 音声ドキュメント検索, 講演音声, パッセージ検索, ドキュメント拡張

Spoken Document Retrieval for Oral Presentations Integrating Global Document Similarities into Local Document Similarities

HIROAKI NANJO^{1,a)} YUSUKE IYONAGA¹ TAKEHIKO YOSHIMI¹

Received: September 26, 2011, Accepted: March 2, 2012

Abstract: A spoken document retrieval (SDR) method for oral presentations is addressed. We propose an integration method of global information and local information based on a pseudo topic hierarchy of presentations. Specifically, for detecting a part of oral presentations about 30 seconds to 1 minute (local document), which is simply divided according to the number of utterances, we integrate a similarity between a given query and a whole presentation (global document) into a similarity between the given query and a local document contained in the global document. For the 15-utterances-based, 30-utterances-based and 60-utterances-based pseudo passage retrieval tasks of CSJ SDR, we confirmed a statistical improvements of information retrieval performance (11 point average precision).

Keywords: spoken document retrieval, oral presentation, passage retrieval, document expansion

1. はじめに

講演音声集合から探したい内容を表す箇所を見つける「講演音声ドキュメント検索」の研究を行う. これまで音声ドキュメント検索の主な研究対象は TREC SDR [1] に代

表されるようにニュースであった. ニュースは通常 1 分程度で自己完結的に作られているため, 各ニュース音声を検索単位とするのが自然であり, ドキュメント拡張などを用いて音声認識誤りに対して頑健にインデキシングを行う方法 [2], [3] や, クエリ拡張 [3], [4] などが主な研究テーマであった.

これに対し, 本研究ではある程度の長さを持つ講演や講義の音声を対象とした音声ドキュメント検索手法を研究す

¹ 龍谷大学理工学部
Faculty of Science and Technology, Ryukoku University,
Otsu, Shiga 520-2194, Japan

a) nanjo@rins.ryukoku.ac.jp

る。講演や講義は通常短くても 10 分以上の長さを持っており、従来のようにこのような音声ドキュメントそのものを検索単位とするのは実用上問題がある。すなわち、適切な検索結果が得られても目的とする情報にダイレクトにアクセスできないという問題がある。たとえばユーザが知りたい内容が講演や講義の一部である場合に、検索結果の講演や講義の中から該当箇所を自ら探さなくてはならないという問題がある。そのため、講演や講義を話題のまとまり(サブトピック)ごとに分割し、それらを検索対象とすることが望ましい。ただし、このような講演や講義のトピック分割自体が難しい問題であるため、本研究ではサブトピックごとへの分割は行わず、講演音声をあらかじめ 30 秒から 1 分程度の単位に機械的に区切っておき、それらの区間を検索対象とすることで、欲しい情報の近くにダイレクトにアクセスできるようにすることを考える。このような、ドキュメントの一部を検索対象とする検索はパッセージ検索とよばれる。本研究はあらかじめ機械的に分割された発話区間をパッセージと見なすパッセージ検索と位置づけられる。

30 秒から 1 分程度の短い講演音声ドキュメントを検索単位、すなわちパッセージと見なした場合、検索対象ドキュメントの総数が増えるうえに、検索結果を絞り込むための文脈情報が不足する。この結果、不適切な検索結果が提示されることがあり、高い検索性能を得るのが難しいという問題がある。これに対して本研究では、講演単位などの大きなまとまり(広域文書)での粗い検索と、30 秒から 1 分程度の単位(局所文書)の検出したい単位での詳細な検索とを統合することを考える。具体的には、クエリと広域文書との類似度を局所文書との類似度に統合して、局所文書を検索する手法を提案する。

局所文書(パッセージ)の検索では、局所文書集合を回答候補として出力しておき、他の知識源を用いて回答候補を再重みづけ(リランキング)することが広く行われており、本提案手法はこの手法と位置づけられる。テキストを対象とする情報検索では、単純なモデル(たとえばベクトル空間モデル)で出力した回答候補に対して、高度な知識を用いた絞り込みや再重みづけが行われている。たとえば、検索クエリの質問タイプを推定して回答を絞ること [5], [6] や種々の QA パターンから回答と質問の対応確率を学習しておき、それをもとに回答を絞ること [7] などが行われている。提案手法は、絞り込みのための高度な知識を必要とせず、自動生成された広域文書と局所文書間の包含関係のみを用いる点で、単純かつ新しい。

提案手法は XML に代表されるような構造化テキストの検索手法 [8], [9] とみることでもできる。ただし、音声ドキュメントではタイトルや段落区切りなどの構造が明示されないため [10]、構造化テキストに対する従来の検索手法は音声ドキュメント検索に利用できない。提案手法は、講演の

ような長い音声ドキュメントに対して、自動的に局所文書および広域文書群を生成し、それらを構造化テキストのように用いる点において新規性を有する。

提案手法は、局所文書を広域文書でドキュメント拡張したものとみることでもできる。ドキュメント拡張とは、検索対象ドキュメントの類似テキストを WEB などから集め、そこに含まれる語をそのドキュメントの索引語として追加する手法である。一般的には、ドキュメントからの WEB 検索質問の生成自体が難しいうえに、選択されたテキストが元のドキュメントと類似しているかを判定することが難しく、効果を得にくいという問題がある [2], [11], [12]。本手法は、局所文書と広域文書の包含関係のみを用いる、新たにテキストを収集することを必要としない、講演の局所文書検索のための効果的なドキュメント拡張手法といえる。本田ら [13] は各発話に対してその直前直後の文集合によるドキュメント拡張を行っている。各発話を局所文書、周辺の文集合を広域文書と考えると提案手法と類似しているものの、本研究と異なり、より広範囲の情報や講演全体の情報、およびそれらの複合的な利用は行っていない。なお、杉本ら [14] は自動分割した局所文書群から広域文書検索のための複数の拡張索引の生成手法を提案しており、これは広域文書を局所文書でドキュメント拡張する手法と見なすことができる。局所文書と広域文書を利用している点で類似性がみられるが、本提案手法は局所文書の検索時に広域文書の情報を利用するものであり、本質的に異なる。

本論文の構成について述べる。2 章では、現在、日本で利用できる大規模な講演音声ドキュメント検索のテストコレクションについて述べ、3 章では、本研究でのベースライン音声ドキュメント検索システムについて述べる。4 章では提案手法について述べ、5 および 6 章では提案手法の有効性を示す。7 章で結論を述べる。

2. 講演音声ドキュメント検索

本研究では、講演音声を対象として検索を行う。講演を記録したものには音声だけでなくスライド画像などが含まれることがある。スライド文字を解析 [15] して音声認識や情報検索に用いることも考えられるが、本研究では、これは扱わず音声のみを検索対象として検索する方法を研究する。

2.1 検索評価用テストコレクション

情報検索システムの評価を行ううえで、クエリに対してドキュメント集合中のどのドキュメントが適合しているかという情報が必要である。テストコレクションとは、ドキュメント集合、クエリ集合、適合情報を備えた情報検索システムの評価用データである。

本研究では、音声ドキュメント処理ワーキンググループによって作成されたテストコレクション [16] を用いて研究

を行う。これは、日本語話し言葉コーパス [17] (以後, CSJ と記す) の学会講演 987 件と模擬講演 1,715 件の合計 2,702 件の講演を検索対象とするものである。学会講演の長さはほとんどが 10 分から 25 分程度であるが、1 時間を超えるものもある。模擬講演は、一般話者による日常的话题についての 12 分程度のスピーチである。テストコレクションには、この 2,702 件の音声に対しての音声認識結果 (認識率は 65% から 95%) も含まれている。本研究は、この音声認識結果を利用して、講演音声ドキュメント検索の実験を行う。クエリは自然言語文で記述された 39 件のテキストであり、各クエリに対する答えとしての適合情報が、どの講演のどの発話からどの発話までという単位で付与されている。なお適合度として適合 (R) と部分適合 (P) が存在する。本研究では適合ラベル (R) が付与された区間をクエリに対する正解として扱った。

このテストコレクションでのクエリは、主に non-factoid 型のクエリである [16]。ただし、「ペットの犬の名前のリストアップ」のような factoid 型のクエリも存在する。本研究では音声ドキュメントをあらかじめ音声認識してテキスト化しておき、そのテキストを検索対象とするため、本テストコレクションでの提案手法はテキストドキュメントを対象とした non-factoid 型の検索手法と見なすこともできる。ただし、段落情報 [6] や明示的なトピック情報が利用できないこと、および音声認識誤りを含むことなどから、従来のテキスト検索のタスクよりも難しいと考えられる。

2.2 評価尺度

情報検索システムの検索性能の評価は、再現率 (recall) と精度 (precision) を用いるのが一般的である。本研究では、評価尺度として式 (1) で示すこれらを組み合わせた評価尺度である 11 点平均精度 (11-point Average Precision, “11ptAP” と記す) [18] を用いる。これは各検索クエリ Q_k に対して 0.0 から 1.0 まで 0.1 刻みでの各再現率レベル x における補間精度 $IP_{Q_k}(x)$ (式 (3)) を求め、それらの平均 $AP(Q_k)$ (式 (2)) を全検索クエリで平均したものである。

$$11ptAP = \frac{1}{N} \sum_{k=1}^N AP(Q_k) \quad (1)$$

$$AP(Q_k) = \frac{1}{11} \sum_{i=0}^{10} IP_{Q_k} \left(\frac{i}{10} \right) \quad (2)$$

$$IP_{Q_k}(x) = \max_{x \leq R_{Q_k}(T)} P_{Q_k}(T) \quad (3)$$

ここで $R_{Q_k}(T)$ と $P_{Q_k}(T)$ は、それぞれクエリ Q_k に対する検索結果の上位 T 番目までの検索結果の再現率と精度である。

今回は、1 つのクエリに対して類似度が 0 でないものを全件検索し、検索結果全体での再現率よりも高い再現率レベル x の補間精度 $IP_{Q_k}(x)$ は 0 とした。

3. 検索システム

本研究ではベクトル空間モデル [18] に基づくドキュメント検索システムを用いる。ベクトル空間モデルは、ドキュメントとクエリをベクトルで表現し、ベクトル間の距離により検索を実現するモデルである。本研究では、ベクトル間の類似度に SMART [19] を用いる。すなわち、あるクエリ Q とドキュメント D_i ($1 \leq i \leq N$) の類似度を、 Q と D_i のそれぞれでの索引語 t_k ($1 \leq k \leq m$) の正規化出現頻度 q_{t_k} および d_{i,t_k} を用いて、式 (4) で与えるものである。

$$SMART(Q, D_i) = \sum_{k=1}^m (q_{t_k} \cdot d_{i,t_k}) \quad (4)$$

ただし、

$$d_{i,t_k} = \begin{cases} \frac{1 + \log(\text{tf}_{i,t_k})}{1 + \log(\text{avtf})} & \text{if } \text{tf}_{i,t_k} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$q_{t_k} = \begin{cases} \frac{1 + \log(\text{qtf}_{t_k})}{1 + \log(\text{avqtf})} \log \frac{N}{n_{t_k}} & \text{if } \text{qtf}_{t_k} > 0 \\ 0 & \text{otherwise} \end{cases}$$

ここで、 tf_{i,t_k} は D_i 中での t_k の出現数、 avtf は D_i における単語の出現数の平均を表す。pivot は 1 ドキュメント中の異なり単語数の平均、 utf_i は D_i 中の異なり単語数を表す。slope は補間係数 (0.2) である。 qtf_{t_k} は、 Q 中での t_k の出現数、 avqtf は Q に含まれる単語の出現数の平均を表す。 N は検索対象のドキュメント集合の全ドキュメント数を表す。 n_{t_k} は、 t_k を含むドキュメントの数を表す。

本研究では、クエリ Q が与えられたとき、すべてのドキュメント D_i について Q との類似度 $SMART(Q, D_i)$ を算出し、類似度が 0 より大きいものを高い順に全件出力する。

4. 講演音声ドキュメント検索における広域文書を用いた局所文書の検索手法

本研究では、講演には構造があり、探したい部分が講演のトピックのサブトピックとなっている可能性が高いという点に着目し、講演音声を検索する方法を提案する。具体的には、探し出したい講演の一部を局所文書、局所文書よりも大きな単位を広域文書と定義し、局所文書のトピックは当該局所文書を包含する広域文書のトピックの一部となっていると見なして、クエリと広域文書との類似度を局所文書との類似度に統合して、局所文書を検索する手法を提案する。この様子を図 1 に示す。局所文書としては、トピック分割を行わずに機械的に分割したもの、具体的には 30 秒から 1 分程度の長さを持ち、秋葉らによるテストコレクションの検索性能の基本評価 [16] でも利用されていた

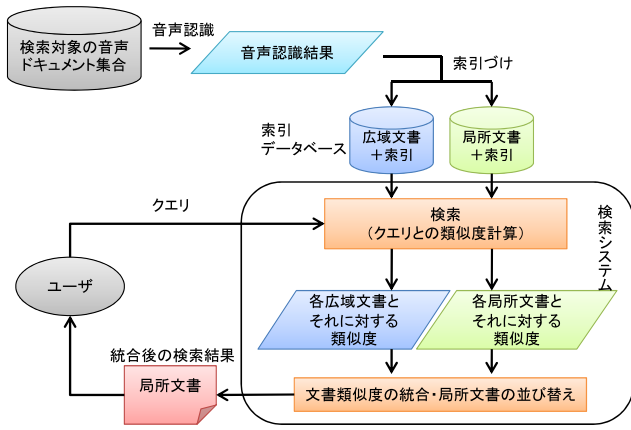


図 1 類似度の統合による音声ドキュメント検索

Fig. 1 Spoken document retrieval for oral presentations based on local and global documents.

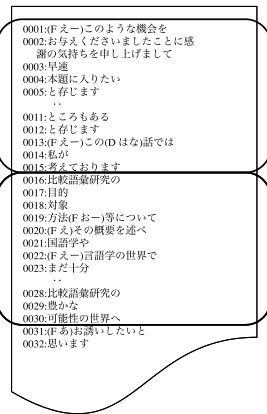


図 2 15 発話区間

Fig. 2 15-utterances-based pseudo passage unit.

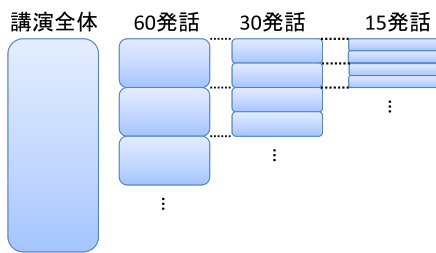


図 3 局所文書 (15 発話単位) と広域文書 (30 発話単位, 60 発話単位, 講演単位) の関係図

Fig. 3 Local document (15-utterances) and global documents (30-utterances, 60-utterances, and a whole presentation).

15 発話*1を採用する。15 発話単位は図 2 に示すように講演の先頭から順に 15 発話ごとに区切ったものであり、この各区間を 1 つの局所文書とする。その際の広域文書としては 15 発話を包含する 30 発話、60 発話、講演全体を採用する (図 3)。本研究では、30 発話と 60 発話についても 15 発話と同様に講演の先頭から順にそれぞれ 30 発話、60

*1 単純に全講演の長さの合計をドキュメント数で割ると、1 ドキュメントあたり 35 秒程度。

発話ごとに区切る。なお、先頭から順に 15, 30, 60 発話で分割を進めた場合に、講演の最終部分で端数が生じることがある。この部分がそれぞれ 15, 30, 60 発話に満たない場合でも、それを 1 つのドキュメントとして扱う。CSJ 講演音声ドキュメントテストコレクションでのこれらのドキュメント数を表 1 に示す。最終部分で生じる端数の影響により、60 発話単位と 30 発話単位、30 発話単位と 15 発話単位でそれぞれドキュメント数が倍にはならない。また、最終部分の端数の箇所を除き、1 つの 30 発話単位ドキュメントはそれに対応する連続する 2 つの 15 発話単位ドキュメントを結合したものと同一であり、1 つの 60 発話単位ドキュメントは連続する 2 つの 30 発話単位ドキュメントを結合したものと同一である。なお、局所文書および広域文書の定義の方法は任意であり、様々なものを考えることができる。本研究では、テストコレクションでの基本的な検索性能 [16] との比較ができるよう、局所文書としてこのような単位を選択した。本論文での広域文書の定義方法は一実装であり、他の定義方法の検討の余地はある。

本論文では、広域文書と局所文書類似度の統合を行わず、局所文書を検索する方法をベースラインの手法とし、広域文書類似度と局所文書類似度の統合手法 (提案手法) の有効性を示す。また、局所文書がある程度の長さを持つ場合、具体的には局所文書を 30 発話単位とした場合および 60 発話単位とした場合に、広域文書情報を利用する効果についても考察を行う。

4.1 1 種類の広域文書の利用

はじめに、クエリと 1 種類の広域文書類似度を、クエリと局所文書類似度に統合する方法を述べる。ここでは局所文書を 15 発話単位として説明を行う。

講演 D_i の l 発話目から $l+n$ 発話目までの発話区間を $D_{i,l}^{l+n}$ と表すことにすると、クエリ Q と局所文書 (15 発話単位) との類似度は $\text{SMART}(Q, D_{i,k}^{k+14})$ と表現できる。局所文書を包含する広域文書とクエリとの類似度を統合した後の、局所文書とクエリの類似度 $S(Q, D_{i,k}^{k+14})$ は、 Q と広域文書 $D_{i,l}^{l+n}$ の類似度 $\text{SMART}(Q, D_{i,l}^{l+n})$ と、 Q と $D_{i,k}^{k+14}$ との類似度 $\text{SMART}(Q, D_{i,k}^{k+14})$ を対数線形補間したものと定義する (式 (5))。なお、対数線形補間では、一方の類似度が高くても他方の類似度が低い場合は最終的な類似度が低くなるため、両類似度が高いドキュメントのみが選択される。

$$S(Q, D_{i,k}^{k+14}) = (1 - \lambda) \log \text{SMART}(Q, D_{i,l}^{l+n}) + \lambda \log \text{SMART}(Q, D_{i,k}^{k+14}) \quad (5)$$

ここで、 $D_{i,l}^{l+n}$ は $D_{i,k}^{k+14}$ を包含しているため、 $l \leq k$, $k + 14 \leq l + n$ の条件を満たす。 λ は各類似度に対する重み係数であり、0 から 1 までの値をとるものとする。

表 1 CSJ 講演音声ドキュメント検索テストコレクションでの局所文書 (15 発話単位) および 広域文書 (30 発話単位, 60 発話単位, 講演単位) のドキュメント数

Table 1 Numbers of local documents (15-utterances) and global documents (30-utterances, 60-utterances, and whole presentations) on the spoken document retrieval test collection of CSJ.

	15 発話単位	30 発話単位	60 発話単位	講演単位
ドキュメント数	60,202	30,762	16,060	2,702

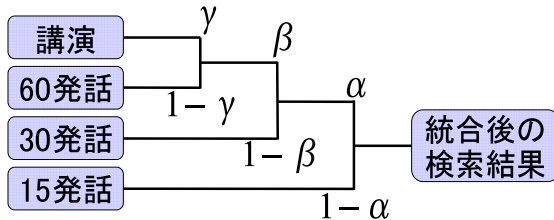


図 4 複数の広域文書類似度の階層的統合による講演音声ドキュメント検索

Fig. 4 Spoken document retrieval for oral presentations based on hierarchical integration of global document similarities.

4.2 複数の広域文書類似度の統合

次に、複数の広域文書を用いる手法について述べる。具体的にはクエリと講演単位, 60 発話単位, 30 発話単位の 3 種類の広域文書の類似度を, 15 発話単位との類似度に統合する方法について述べる。

本研究では、包含関係にある各広域文書を大きさの順に並べ、段階的に統合を行う。具体的には、式 (6) に示すように、まず、講演単位の広域文書との類似度と 60 発話単位の広域文書との類似度を重み γ を用いて統合し、次に、その類似度と 30 発話単位の広域文書との類似度を重み β を用いて統合する。最後に、局所文書 (15 発話単位) との類似度を重み α を用いて統合し、最終的な結果とする。この様子は、図 4 にも示されている。

$$\begin{aligned}
 S(Q, D_{i,k}^{k+l4}) &= \left(1 - \alpha\right) \log \text{SMART}(Q, D_{i,k}^{k+l4}) \\
 &+ \alpha \left(\left(1 - \beta\right) \log \text{SMART}(Q, D_{i,l}^{l+29}) \right. \\
 &+ \beta \left(\left(1 - \gamma\right) \log \text{SMART}(Q, D_{i,m}^{m+59}) \right. \\
 &\left. \left. + \gamma \log \text{SMART}(Q, D_i) \right) \right) \quad (6)
 \end{aligned}$$

ここで、各広域文書はそれよりも小さい単位の広域文書/局所文書を包含しているため、それぞれ $m \leq l \leq k$, $k + 14 \leq l + 29 \leq m + 59$ を満たすものである。また、統合重み α, β, γ はそれぞれ 0 から 1 の間の値をとるものとする。

5. 評価実験

5.1 評価方法

本提案手法は、自動分割により得られた局所文書を検索対象とするものである。本論文での実験では、各クエリに対する正解ドキュメントを、テストコレクションで定義された正解発話区間 (適合ラベル (R) がつけられた区間) が一部でも含まれるすべての局所文書とした。なお、適合ラベル (R) がついている発話区間の長さの分布は図 5 のとおりであり、平均長は 2.76 発話であった。ただし、講演全体に適合ラベル (R) を付与しているものが 7 つあり、それらの長さ (講演長) は分布と平均長の算出からは除いている。

5.2 ベースラインシステムによる検索結果

本論文では、先行研究 [20], [21] に基づき、索引語を形態素の基本形 (名詞, 動詞のみ) とし、検索エンジンには汎用連想計算エンジン GETA [22] を用いてベースライン検索システムを設計した。

ベースラインシステムを用いて、CSJ の 2,702 講演を 15 発話ごとに区切り各区間を検索対象のドキュメントとして検索を行った。表 1 に示すとおり、ドキュメント数は 60,202 である。ベースラインシステムでの 11 点平均精度は 0.177 であった。

5.3 提案手法による検索結果

5.3.1 1 種類の広域文書類似度の統合

はじめに、4.1 節で述べた、クエリと 1 種類の広域文書の類似度を、クエリと局所文書の類似度に統合する方法による検索結果について述べる。

結果を表 2 に示す。ここでは、39 件のクエリを 13 のセットにわけ、交差検定を行った。具体的には、36 件のクエリを用いて検索精度が最大となるように式 (5) の統合重み λ の推定を行い、残りの 3 件の検索実験にその推定重みを使用した。これを 13 の分割セットそれぞれに対して行った。なお、講演単位を広域文書とした場合の統合重み λ は、どの分割セットでも 0.48 であり、広域文書を 60 発話単位としたときは、統合重みはすべて 0.20 であった。広域文書を 30 発話単位としたときは、統合重みは 0.08 から 0.17 であった。ただし、8 セットに対して 0.14, 3 セット

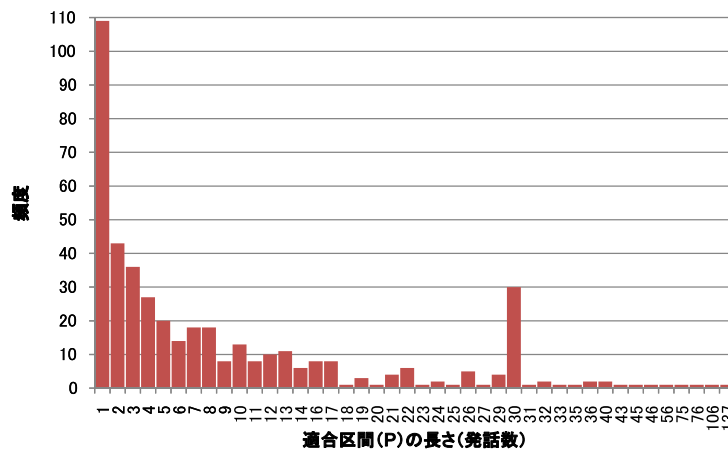


図 5 テストコレクションでの適合区間 (R) の長さ (発話数) の分布

Fig. 5 Distribution of length (number of utterances) of relevance segments (R) in the Japanese SDR test collection.

表 2 1 種類の広域文書類似度の統合

Table 2 Integration of one global document similarity into local document similarity.

局所文書	広域文書	11 点平均精度
15 発話	なし (ベースライン)	0.177
	30 発話	0.222
	60 発話	0.231
	講演全体	0.226

表 3 複数の広域文書類似度の階層的統合の効果

Table 3 Effect of hierarchical integration of global document similarities.

局所文書	広域文書	11 点平均精度
15 発話	なし (ベースライン)	0.177
	講演全体+60 発話+30 発話	0.241
	講演全体+60 発話 ($\beta = 1$)	0.227
	講演全体+30 発話 ($\gamma = 1$)	0.244
	60 発話+30 発話 ($\gamma = 0$)	0.238

に対して 0.15 であり、分割セット間での統合重みの大きな違いはみられなかった。

広域文書の情報を用いることで、広域文書を用いない検索精度 (0.177) よりも高い精度が得られた。ベースラインからの改善について、対応のある 2 群の差の検定 (T 検定) を行ったところ、どの広域文書を用いたときでも有意水準 1% で有意差が認められた。本実験では広域文書を 60 発話単位としたときに検索精度が最も高くなり、0.231 が得られた。ただし、他の広域文書 (30 発話、講演全体) を使用した場合の検索精度との有意な差は確認できなかった。

5.3.2 複数の広域文書類似度の統合

次に、4.2 節で述べた、複数の広域文書の統合による検索 (階層的検索) の結果について述べる。具体的には講演単位、60 発話単位、30 発話単位の 3 種類の広域文書とクエリとの類似度を、15 発話単位とクエリとの類似度に統合する方法による結果について述べる。

結果を表 3 (広域文書：講演全体+60 発話+30 発話) に示す。広域文書を 1 つだけ使った場合 (表 2) よりも高い検索精度 0.241 が得られた。なお、式 (6) の統合重み γ は 0.77 から 0.84 の間、 β は 0.24 から 0.34 の間、 α は 0.10 から 0.17 の間の値であり、各分割セット間での大きな違いはみられなかった。

ベースラインおよび広域文書として 60 発話単位文書のみを使った場合の検索結果と比べて、階層的検索の結果が

有意に高いかについて、対応のある 2 群の差の検定 (T 検定) を行ったところ、有意水準 1% で有意差が認められた。このことは、クエリと広域文書との類似度を局所文書との類似度に統合する提案手法は有効であることを示している。

比較のため、広域文書を 2 種類にした場合の結果も表 3 に示してある。これらは、それぞれ統合重み β を 1 に固定、 γ を 1, 0 に固定して階層的統合したことに相当する。広域文書を講演全体+60 発話とした場合および 60 発話+30 発話とした場合は、広域文書として 60 発話単位のみを使用した場合と有意な差がみられなかった。広域文書を講演全体+30 発話とした場合は、広域文書として 60 発話単位のみを使用した場合の検索結果と比べて、有意に精度が高かった (有意水準 1%)。また、広域文書を講演全体+60 発話+30 発話とする場合と講演全体+30 発話とする場合とでは、有意な差はみられなかった。60 発話単位は講演全体および 30 発話単位の中間的な長さであるためどちらの特徴も含んでいると考えられ、60 発話単位を階層的統合に用いる効果が得られなかったと考えられる。

図 6 にクエリごとの検索精度を示す。Akiba ら [16] の実験結果と同様に、クエリごとに検索精度にばらつきがみられる。1 つの広域文書を用いた 3 通りの場合を比較すると、クエリごとに、長い/短い広域文書を用いる効果が高い場合と低い場合とがあることが分かる。これらの 3 つの広

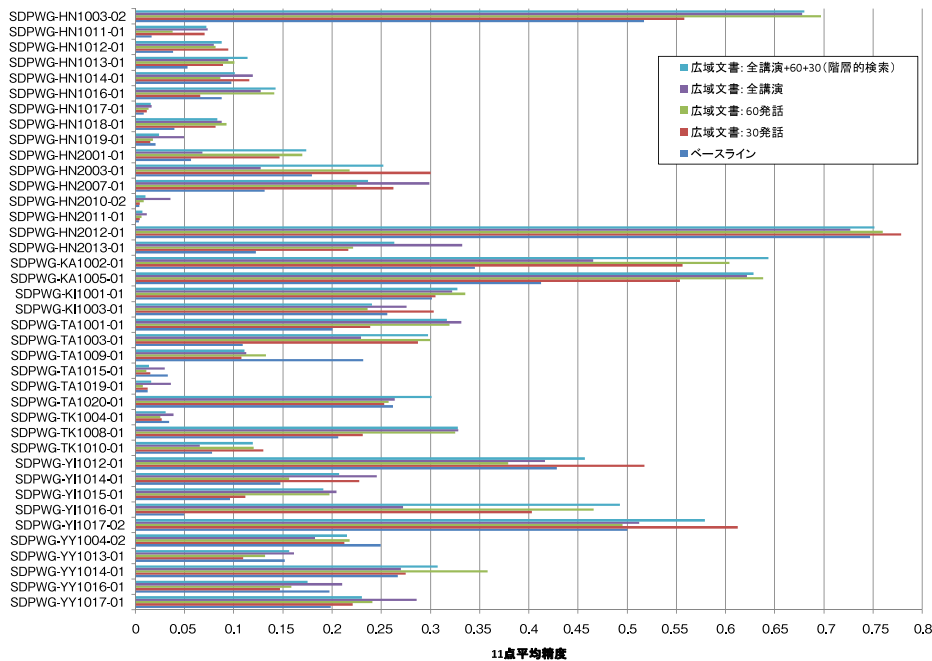


図 6 クエリごとの 11 点平均精度

Fig. 6 11ptAP for each query.

域文書を用いた検索（階層的検索）によって1つの広域文書のみを用いた場合よりも高い精度が得られたクエリは39件中6件、低い精度となったクエリは0件であった。すなわち、階層的検索では1種類の広域文書を用いた場合の最低の検索精度よりも低い検索精度となることはなかった。これらのことより、階層的統合は、ある長さの広域文書を1種類だけ使う場合に比べて、クエリの変更に対して頑健であると考えられる。

5.4 局所文書の長さと同域文書利用の効果

次に、局所文書が長い場合の、広域文書を利用した検索の結果について述べる。

はじめに、局所文書を30発話単位とした場合について述べる。30発話単位に基づく局所文書の検索タスクはCSJ講演音声ドキュメント検索テストコレクションでは30,762ドキュメントの検索タスクとなる(表1)。ベースラインシステムで検索を行ったところ、検索精度(11点平均精度)は0.255であった。

広域文書として60発話単位と講演単位を用いて、クエリと広域文書の類似度を局所文書との類似度に統合して検索を行った。結果を表4に示す。すべての手法でベースラインからの改善が得られ、これらの改善は有意であった(有意水準5%)。広域文書として講演単位を利用したときに、最も高い検索精度0.283が得られた。広域文書を60発話単位とした場合は改善幅は小さかった。60発話単位と講演単位の音声ドキュメントの階層的利用でも、改善幅は小さかった。これらの理由として、30発話単位の音声ドキュメントはやや長く*2、その発話前後の情報を十分に含んでお

表 4 局所文書を30発話とした場合の検索性能

Table 4 Information retrieval performance for 30-utterances retrieval task.

局所文書	広域文書	11点平均精度
30 発話	なし (ベースライン)	0.255
	60 発話	0.266
	講演全体	0.283
	講演全体+60 発話	0.276

表 5 局所文書を60発話とした場合の検索性能

Table 5 Information retrieval performance for 60-utterances retrieval task.

局所文書	広域文書	11点平均精度
60 発話	なし (ベースライン)	0.310
	講演全体	0.320

り、60発話単位の広域文書を用いる効果が低かったことが考えられる。講演単位の広域文書を用いた場合では、近傍以外の部分の情報をうまく取り入れられたと考えられる。

次に、局所文書を60発話単位とした場合の結果について述べる。60発話単位に基づく局所文書の検索タスクは、テストコレクションでは16,060文書の検索タスクとなる(表1)。ベースラインシステムで検索を行ったところ、検索精度(11点平均精度)は0.310であった。広域文書として、講演単位を利用した。結果を表5に示す。検索性能の有意な改善(5%水準)はみられたものの、局所文書が短い場合(15発話や30発話)に比べた場合、向上する割合

*2 単純に全講演の長さの合計をドキュメント数で割ると、1ドキュメントあたり1分10秒程度。

が小さくなっていることが分かる。局所文書が長くなるに従って、検索に必要な索引語が多く得られていく結果と考えられる。

これらのことは、局所文書を長くするに従って広域文書類似度を統合する効果は次第に小さくなるものの、広域文書の利用は局所文書の検索に効果的であることを示している。提案手法は、講演に代表される、長くかつ話題に階層構造を持つ音声ドキュメントの一部の短い区間を検索するタスクにおいて有効であると考えられる。

なお、本論文での広域文書の定義方法は一実装であり、最適な広域文書の単位（定義方法）の研究は今後の課題である。

6. 講演の人手書き起こしドキュメントに対する検索

これまで、講演音声の自動音声認識結果に対して提案手法の有効性を示した。次に、音声認識誤りを含まない講演の人手書き起こし（議事録作成の第1段階）に対して本手法を適用した結果を示す。結果を表6に示す。音声認識結果に対する場合と同様に、提案手法により有意な改善がみられた。

人手書き起こしドキュメントに対する検索（表6）と音声認識結果に対する検索の実験結果（表2および表3）とを比較すると、音声認識結果に対する検索において提案手法による改善が大きくなっている傾向、すなわち提案手法の音声認識誤りに対する頑健性が確認できた。ベースラインの検索手法では、音声認識誤りによる検索精度の低下は0.018ポイント（0.195→0.177）であったのに対し、階層的検索では音声認識誤りによる検索精度の低下は0.012ポイント（0.253→0.241）であった。

短い局所文書には含まれる索引語が少ないために、各索引語の認識誤りの影響が大きく、特に検索にとって重要な語が音声認識誤りされた場合には、それを含む局所文書とクエリとの類似度が低く算出されてしまう。逆に、誤って重要な語が索引語に追加された場合は、局所文書とクエリとの類似度を大きく算出してしまふ。これに対して、広域文書では、重要な語の一部に音声認識誤りが存在しても、他の部分には正しく音声認識された重要な語が存在する可能性

表6 書き起こしドキュメントに対する広域文書類似度統合の効果
Table 6 Effect of integration of document similarities for correct transcriptions.

局所文書	広域文書	11点平均精度
15発話	なし（ベースライン）	0.195
	30発話	0.239
	60発話	0.242
	講演全体	0.235
	講演全体+60発話+30発話	0.253

が高く、広域文書とクエリとの類似度が、誤って小さくまたは大きく見積もられる現象は起きにくいと考えられる。すなわち、音声認識結果に対する検索では、局所文書中の単語の認識誤りにより検索性能が低下する問題が大きかったが、提案手法ではそのような局所的な認識誤りによる検索への悪影響を、広域文書を用いることで抑制できていると考えられる。

7. おわりに

講演音声ドキュメント検索において、短い発話区間（局所文書）と、局所文書を包含する広域文書を用いて局所文書を検索する手法を提案し、CSJを対象とした音声ドキュメント検索テストコレクションで評価実験を行った。局所文書を30秒から1分程度の15発話単位とした場合は、広域文書を用いない従来の検索方法での精度が0.177であったのに対し、広域文書を用いる提案手法により有意な精度の改善が得られ、検索精度0.241が得られた。また、局所文書の長さを30発話、60発話と長くしたときにも、局所発話が短い場合（15発話）に比べると広域文書類似度を統合する効果は小さくなるものの、局所文書の検索に効果的であることが分かった。講演の書き起こしテキスト（音声認識誤りなし）に対しても提案手法は有効であることが分かり、音声認識結果（本タスクでは音声認識率65%から95%）に対して本手法は、より有効に機能することも分かった。

参考文献

- [1] Voorhees, E.M. and Harman, D.: Overview of the Ninth Text REtrieval Conference (TREC-9), *TREC* (2000).
- [2] 杉本樹世貴, 西崎博光, 関口芳廣: 音声ドキュメント検索における Web ページを用いたドキュメント拡張の効果, 情報処理学会研究報告, 2009-SLP-76, No.11 (2009).
- [3] Johnson, S.E., Jourlin, P., Jones, K.S. and Woodland, P.C.: Spoken Document Retrieval for TREC-9 at Cambridge University, *TREC* (2000).
- [4] Gauvain, J.-L., Lamel, L., Barras, C., Adda, G. and de Kercadio, Y.: The LIMSI SDR System for TREC-9, *TREC* (2000).
- [5] 佐々木裕, 磯崎秀樹, 鈴木 潤, 国領弘治, 平尾 努, 賀沢 秀人, 前田英作: SVMを用いた学習型質問応答システム SAIQA-II (自然言語), 情報処理学会論文誌, Vol.45, No.2, pp.635-646 (2004).
- [6] 水野淳太, 秋葉友良: 任意の回答を対象とした質問応答のための実世界質問の分析と回答タイプ判定法の検討, 言語処理学会 第13回年次大会講演論文集, pp.1002-1005 (2007).
- [7] Surdeanu, M., Ciaramita, M. and Zaragoza, H.: Learning to rank answers on large online QA collections, *Proc. 46th Annual Meeting for the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, pp.719-727 (2008).
- [8] 江口浩二: 情報検索のための確率的言語モデルに関する動向と課題, 電子情報通信学会論文誌, Vol.J93-D, No.3, pp.157-169 (2010).
- [9] Ogilvie, P. and Callan, J.: Combining document rep-

presentations for known-item search, *Annual ACM Conference on Research and Development in Information Retrieval*, pp.143-150 (2003).

- [10] Kawahara, T., Hasegawa, M., Shitaoka, K., Kitade, T. and Nanjo, H.: Automatic indexing of lecture presentations using unsupervised learning of presumed discourse markers, *IEEE Trans. Speech & Audio Process*, Vol.12, No.4, pp.409-419 (2004).
- [11] 宇野 有, 伊藤 仁, 伊藤彰則, 牧野正三: 音声ドキュメント検索のための WWW を用いたインデクス改善, 第 4 回音声ドキュメント処理ワークショップ講演論文集, SDPWS2010-09 (2010).
- [12] 金寺 登, 船田哲男, 中川聖一: 検索キーワード補完による講義サブピック検索, 第 4 回音声ドキュメント処理ワークショップ講演論文集, SDPWS2010-10 (2010).
- [13] 本田耕一郎, 秋葉友良: CSJ 音声ドキュメント検索テストコレクションを対象とした部分音声区間の内容検索タスクの設定とその検索手法の検討, 第 4 回音声ドキュメント処理ワークショップ講演論文集, SDPWS2010-15 (2010).
- [14] 杉本樹世貴, 西崎博光, 関口芳廣: 音声ドキュメント検索における音声ドキュメント分割を利用した Web 収集の効果, 日本音響学会 2011 年春季研究発表会講演論文集, pp.121-124 (2011).
- [15] 河原達也, 根本雄介, 勝丸徳浩, 秋田祐哉: スライド情報を用いた言語モデル適応による講義の音声認識と字幕付与, 情報処理学会論文誌, Vol.50, No.2, pp.469-476 (2009).
- [16] Akiba, T., Aikawa, K., Itoh, Y., Kawahara, T., Nanjo, H., Nishizaki, H., Yasuda, N., Yamashita, Y. and Itou, K.: Construction of a test collection for spoken document retrieval from lecture audio data, *IPSJ-Journal*, Vol.50, No.2, pp.501-513 (2009).
- [17] 前川喜久雄: 言語研究における自発音声, 日本音響学会研究発表会講演論文集 (春季), pp.19-22 (2001).
- [18] 北 研二, 津田和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版株式会社, ISBN4-320-12036-1 (2002).
- [19] 小作浩美, 内山将夫, 井佐原均, 河野恭之, 木戸出正継: WWW 検索における複数検索結果の結合処理とその評価, 情報処理学会論文誌, Vol.44, No.SIG 8 (TOD 18), pp.78-91 (2003).
- [20] 重安幸治, 南條浩輝, 吉見毅彦: 日本語講演音声ドキュメント検索における索引付けの検討, 情報処理学会研究報告, SLP-76-8 (2009).
- [21] Shigeyasu, K., Nanjo, H. and Yoshimi, T.: A Study of Indexing Units for Japanese Spoken Document Retrieval, *10th Western Pacific Acoustics Conference (WESPAC X)* (2009).
- [22] 汎用連想計算エンジン GETA, 入手先 (<http://geta.ex.nii.ac.jp>).



弥永 裕介

2010 年龍谷大学理工学部情報メディア学科卒業。在学中に音声ドキュメント検索に関する研究に取り組む。



吉見 毅彦 (正会員)

1987 年電気通信大学大学院計算機科学専攻修士課程修了。1999 年神戸大学大学院自然科学研究科博士課程修了。(財)計量計画研究所(非常勤), シャープ(株)を経て, 2003 年より龍谷大学理工学部勤務。2004 年より 2008 年まで(独)情報通信研究機構専攻研究員を兼任。



南條 浩輝 (正会員)

1999 年京都大学工学部情報学科卒業。2001 年同大学院情報学研究科修士課程修了。2004 年同大学院情報学研究科博士後期課程修了。同年龍谷大学理工学部助手。2007 年同助教。音声言語処理, 特に音声認識・理解に関する研究に従事。日本音響学会, 電子情報通信学会, IEEE, ISCA 各会員。2008 年度日本音響学会栗屋潔学術奨励賞受賞。