講演音声ドキュメント検索のための広域文書類似度と局所文書類似度の統合

弥永 裕介　　南條 浩輝　　吉見 毅彦

龍谷大学 理工学部 情報メディア学科

e-mail: iyonaga@nlp.i.ryukoku.ac.jp

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 に代表されるようにニュースであった．本研究での対象である講演とニュースではドキュメントとしての性質が大きく異なる．ニュースは通常 30 秒から 1 分程度で自己完結的に作られており，各ニュース音声を検索単位とするのが自然であった．これに対し，講演は通常短くても 10 分以上の長さを持っており，このような講演を検索単位とするのは目的とする情報にダイレクトにアクセスできないため不便である．ユーザの利便性を考えると，講演を DVD のチャプタのように 30 秒から 1 分程度に区切り，各区間を検索対象とすることが必要である．しかし，このように短いドキュメントを検索単位とすると一般的に高い検索精度を得るのが難しい．

この問題に対して，我々は講演には講演自体が扱うトピックがあり，さらにその一部となるサブトピックに分けられるという構造があること [1] に着目し，講演単位（広域文書）で絞込んだ上で 1 分程度の単位（局所文書）を検索することを考える．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．

2　音声ドキュメント検索

2.1　検索システム

本研究ではベクトル空間モデル [2] に基づく文書検索システムを用いる．ベクトル空間モデルは，文書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[3] を用いる．すなわち，あるクエリ Q と文書 $D_i(1 \leq i \leq N)$ の類似度を，索引語を $t_k(1 \leq k \leq m)$ として，式（1）で与えるものである．

$$\mathrm{SMART}(Q, D_i) = \sum_{k=1}^{m} (q_{t_k} \times d_{i,t_k}) \qquad (1)$$

ただし，

$$d_{i,t_k} = \begin{cases} \dfrac{1 + \log(\mathrm{tf}_{i,t_k})}{1 + \log(\mathrm{avtf})} \Big/ (1 - \mathrm{slope}) \cdot \mathrm{pivot} + \mathrm{slope} \cdot \mathrm{utf}_i & \text{if } t_k > 0 \\ 0 & \text{otherwise} \end{cases}$$

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

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 文書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 は検索対象の文書集合の全文書数を表す．n_{t_k} は，t_k を含む文書の数を表す．

本研究では，クエリ Q が与えられたとき，全ての文書 D_i について Q との類似度 $\mathrm{SMART}(Q, D_i)$ 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．

2.2　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話し手の身振り手振りや表情，スライドの画像などが含まれることがある．スライドの文字を解析して，索引語に追加す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．

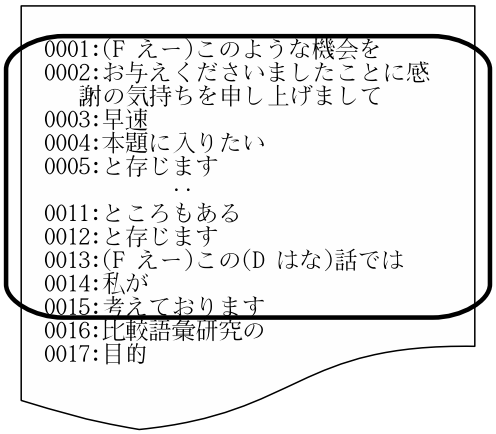

図 1: 15 発話区間

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一部のサブトピックとなっている点に着目し，講演音声を検索する方法を提案する．具体的には，大きなトピック単位を広域文書，探し出したいサブトピック単位を局所文書と定義し，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．局所文書として，1 分程度の長さを持ち，秋葉ら [4] の検索性能評価でも利用されていた 15 発話を採用する．具体的には図 1 に示すように講演の先頭から順に 15 発話ごとに区切り，各区間を 1 つの局所文書とする．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．広域文書と局所文書類似度の統合を行わず，局所文書（15 発話単位）をドキュメントとみなして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．提案手法については 3 章で説明する．

3 広域文書類似度と局所文書類似度の統合

3.1 2 つの文書類似度の統合

はじめに，広域文書の文書類似度と局所文書の文書類似度の 2 者を統合する方法を述べる．$D_{i,l}^{\ l+n}$ を講演 D_i の l から l+n までの発話区間，$D_{i,k}^{\ k+15}$ を講演 D_i の k から k+15 番目までの発話区間とし，統合後のクエリ Q と局所文書 $D_{i,k}^{\ k+15}$ の類似度 $\mathrm{S}(Q, D_{i,k}^{\ k+15})$ は，Q と広域文書 $D_{i,l}^{\ l+n}$ の類似度 $\mathrm{SMART}(Q, D_{i,l}^{\ l+n})$ と Q と $D_{i,k}^{\ k+15}$ との類似度 $\mathrm{SMART}(Q, D_{i,k}^{\ k+15})$ を線形補間したものと定義する（式（2））．ただし，$D_{i,l}^{\ l+n}$ は $D_{i,k}^{\ k+15}$ を包含している，すなわち l≤k，k+15≤l+n をみたすものとする．

$$\begin{aligned}\mathrm{S}(Q, D_{i,k}^{\ k+15}) &= (1-\lambda)\mathrm{SMART}(Q, D_{i,l}^{\ l+n}) \\ &\quad +\lambda\mathrm{SMART}(Q, D_{i,k}^{\ k+15}) \quad (2)\end{aligned}$$

λ は各文書類似度に対する重み係数であり，λ を 0.0 から 1.0 まで 0.1 刻みで変化させて実験を行う．

また，文書類似度を対数線形補間する方法も用いる（式（3））．

$$\begin{aligned}\mathrm{S}(Q, D_{i,k}^{\ k+15}) &= (1-\lambda)\log\mathrm{SMART}(Q, D_{i,l}^{\ l+n}) \\ &\quad +\lambda\log\mathrm{SMART}(Q, D_{i,k}^{\ k+15}) \quad (3)\end{aligned}$$

対数線形補間では，一方の文書類似度が高くても，他方の文書類似度が低い場合は最終的な文書類似度が低くなるため，両類似度が高い文書のみが選択される．

3.2 4 つの文書類似度の統合

次に，講演単位，60 発話単位，30 発話単位，15 発話単位の 4 者の文書類似度を統合する方法を式（4）に示す．まず，講演単位との類似度と 60 発話との類似度を重み γ を用いて統合する．次に，統合した結果と 30 発話との類似度を重み β を用いて統合する．同様に，15 発話との類似度を重み α を用いて統合し，最終的な結果とする．

$$\begin{aligned}\mathrm{S}(Q, D_{i,k}^{\ k+l5}) &= \Big(1-\alpha\Big)\mathrm{SMART}(Q, D_{i,k}^{\ k+l5}) \\ &\quad +\alpha\Bigg(\Big(1-\beta\Big)\mathrm{SMART}(Q, D_{i,l}^{\ l+30}) \\ &\quad +\beta\Big((1-\gamma)\mathrm{SMART}(Q, D_{i,m}^{\ m+60}) \\ &\quad +\gamma\mathrm{SMART}(Q, D_i)\Big)\Bigg) \quad (4)\end{aligned}$$

ただし，l≤k，m≤k，k+15≤l+30，k+15≤m+60 である．これについても対数線形補間を検討する．

4 評価実験と考察

4.1 検索対象と評価尺度

4.1.1 検索評価用テストコレクション

本研究では，音声ドキュメント検索処理 WG によって作成されたテストコレクション [4] を用いて研究を

行った．以下にこの詳細について述べる．

● 文書集合

検索対象の文書集合は，日本語話し言葉コーパス（CSJ）[5] の学会講演 987 件と模擬講演 1715 件の合計 2702 件である．模擬講演は，一般話者による日常的話題についての 12 分程度のスピーチである．テストコレクションでは，この 2702 件の音声に対して音声認識が行われており，認識率は 65 %から 95 %である．本研究では，この音声認識結果の 1-best を使って研究を行う．

● クエリ集合

テストコレクションにおける検索システムの評価用のクエリ集合は，クエリの性質と講演音声を対象とすることを考慮して，以下をみたす 39 件のクエリとなっている．

- 内容を問うクエリ
- 10 件程度の適合情報が存在する
- 特定の分野に偏りがないクエリ

● 適合情報

テストコレクションには検索質問集合の各検索質問文に対し，適合，部分適合，不適合の 3 段階が用意されており，本研究では適合判定のみを用いて評価する．

4.1.2 評価尺度

検索性能の評価尺度には，式（5）で示す補間 11 点平均精度（Interpolated 11-points Average Precision,"11ptAP"と記す）を用いる．これは各検索クエリ Q 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 $IP_Q(x)$ を求め，それらの平均 $AP(Q)$ を全検索クエリで平均をとったものである．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 $IP_Q(x)$ は 0 とした．

$$11\text{pt}AP = \frac{1}{N}\sum_{k=1}^{N} AP(Q) \tag{5}$$

$$AP(Q) = \frac{1}{11}\sum_{i=0}^{10} IP_Q(\frac{i}{10})$$

表 1: 15 発話単位での検索性能

索引語	11ptAP
形態素の出現形	0.152
形態素の基本形 (名詞，動詞のみ)	0.186

$$IP_Q(x) = \max_{x \le R_{qi}} P_{qi}$$

ここで R_{qi} と P_{qi} は，それぞれクエリ Q に対する検索結果の上位 i 番目までの検索結果の再現率と適合率である．

4.2 ベースラインシステムによる検索結果

CSJ の 2702 講演を索引付けして検索を行った．2702 講演を 15 発話ごとに区切り各区間を検索対象の文書とした場合，文書数は 60202 となる．索引語には [6][7][8] を参考に形態素の出現形，形態素の基本形（名詞，動詞のみ）の 2 種類，検索システムには汎用連想計算エンジン GETA[9] を用いた．

実験結果を表 1 に示す．索引語に形態素出現形を用いた場合の 11 点平均精度は 0.152 であった．索引語に形態素基本形（名詞，動詞のみ）を用いた場合に 11 点平均精度は 0.186 であった．このことは索引語として形態素基本形（名詞，動詞のみ）を用いることが有効であることを示しており，この結果は [6][7][8] の傾向と一致する．以後の評価実験においては形態素基本形（名詞，動詞のみ）を索引語とする．

4.3 広域文書類似度と局所文書類似度の統合による検索

4.3.1 2 種類の文書類似度の統合

広域文書類似度と局所文書類似度の 2 者の統合の結果について述べる．局所文書を 15 発話単位とし，広域文書として講演単位，60 発話単位，30 発話単位の 3 通りを試した．実験は 39 クエリに対し 3 分割交差検定で行った．使用した重み λ の具体的な値は以下の通りである．広域文書を講演単位とし，線形補間により類似度の統合を行った場合の重み λ は，3 通りとも 0.7 であり，対数線形補間の場合は 0.42，0.46，0.46 であった．広域文書を 60 発話とした場合の重み λ は 0.12，0.07，0.17（線形補間），0.28，0.1，0.29（対数線形補間）であった．広域文書を 30 発話とした場合の重み λ は 0.2，0.03，0.2（線形補間），0.17，0.04，

表 3: 4 種類の文書類似度の統合

局所文書	広域文書	類似度の統合	11ptAP
15 発話	講演+60 発話+30 発話	線形補間	0.226
		対数線形補間	0.233

表 2: 2 種類の文書類似度の統合

局所文書	広域文書	類似度の統合	11ptAP
15 発話	講演	線形補間	0.232
		対数線形補間	0.228
15 発話	60 発話	線形補間	0.228
		対数線形補間	0.221
15 発話	30 発話	線形補間	0.220
		対数線形補間	0.221

018（対数線形補間）であった．

提案手法による検索性能の評価結果を表 2 に示す．広域文書に講演単位を用いた場合の 11 点平均精度は 0.232 であり，広域文書として 60 発話単位，及び 30 発話単位を用いる場合よりも高い結果となった．これは，長いドキュメントを検索単位としたときの検索精度が高いためと考えられる．また，類似度の統合において式（2）と式（3）のどちらを用いても検索性能に大きな違いは見られなかった．

4.3.2 4 種類の文書類似度の統合

次に，15 発話単位の文書類似度に講演単位，60 発話単位，及び 30 発話単位の類似度の統合（式（4））による検索性能について述べる．これについても 3 分割交差検定を行った．線形補間により類似度の統合を行った場合の重みは，γ が 0.32，0.4，0.32，β が 0.1，0.09，0.15，α が 0.12，0.02，0.05 であった．対数線形補間により類似度の統合を行った場合の重みは，γ が 0.75，0.76，0.67，β が 0.44，0.19，0.4，α が 0.07，0.02，0.18 であった．

検索性能を表 3 に示す．類似度の統合方法として対数線形補間を用いた場合に，本研究で最も高い 0.233 となった．この結果がベースラインの結果よりも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．

5 結論

講演音声ドキュメント検索において，1 分程度の発話区間（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 テストコレクションで評価実験を行った．文書類似度の統合を行わない場合の精度は 0.186 であったが，文書類似度の統合を行った場合に精度が 0.233 に向上した．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の有効性を示した．

参考文献

[1] Tatsuta Kawahara, Masahiro Hasegawa, Kazuya Shitaoka, Tasuku Kitade, and Hiroaki Nanjo. Automatic indexing of lecture presentations using unsupervised learning of presumed discourse markers. In *IEEE Trans. Speech & Audio Process*, Vol.12, No.4. pp.409–419, 2004.

[2] 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.

[3] 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）, 情報処理学会論文誌, pp. 78–91, 2003.

[4] Tomoyosi Akiba, Kiyoaki Aikawa, Yoshiaki Itoh, Tatsuya Kawahara, Hiroaki Nanjo, Hiromitsu Nishizaki, Norihito Yasuda, Yoichi Yamashita, and Katunobu Itou. Construction of a test collection for spoken document retrieval from lecture audio data. In *IPSJ-Journal*, Vol.50. No.2, pp.501–513, 2009.

[5] 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.

[6] 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.

[7] 重安幸治, 南條浩輝, 吉見毅彦. 音声ドキュメント検索における音声認識誤りを考慮した種々の索引付けの検討. 第 12 回関西支部若手研究者交流研究発表会, 2009.

[8] Koji Shigeyasu, Hiroaki Nanjo, and Takehiko Yoshimi. A study of indexing units for japanese spoken document retrieval. In *10th Western Pacific Acoustics Conference (WESPAC X)*, 2009.

[9] 汎用連想計算エンジン GETA. http://geta.ex.nii.ac.jp.