

講演音声ドキュメント検索のための 擬似適合性フィードバック

○西尾 友宏^{1,a)} 南條 浩輝^{2,b)} 吉見 毅彦^{2,c)}

概要：講演や講義などの長い音声ドキュメントの検索のための擬似適合性フィードバック (PRF: Pseudo Relevance Feedback) について述べる。PRF とは、与えられたクエリ (初期クエリ) で検索を行い、その検索結果から初期クエリに関連する語を抽出して、新たなクエリを作成するクエリ拡張手法である。本論文では、はじめに講演音声やその一部を検索対象とした検索 (パッセージ検索) において PRF の効果を調査し、長い音声ドキュメントの検索では、関連語抽出が難しく PRF が効果的でないことを示す。次に、この問題を解決するために、講演音声をあらかじめ 30 秒程度に機械的に区切っておき、それを対象として初期検索を行って関連語抽出を行う PRF 手法を提案する。提案手法が長い音声ドキュメント検索において従来法よりも適していることを明らかにした上で、拡張したクエリによる検索結果と初期クエリによる検索結果を組み合わせたことの効果を示す。CSJ 音声ドキュメントコレクションの講演検索タスクや比較的長めのパッセージ検索タスクにおいて、平均的な検索精度 (11 点平均精度) の向上、および検索精度が向上するクエリ数の増加が確認できた。

Pseudo Relevance Feedback for Lectures

NISHIO TOMOHIRO^{1,a)} NANJO HIROAKI^{2,b)} YOSHIMI TAKEHIKO^{2,c)}

Abstract: Pseudo relevance feedback (PRF) for Lectures is addressed. PRF is one of well-known query expansion (QE) methods, which extracts original query-related terms from initially retrieved documents obtained by an original query. In this paper, we first study PRF on lecture retrieval and passage (part of lectures) retrieval tasks, and show that PRF does not work well on longer spoken document retrieval task due to the difficulty in extracting truly related terms from longer documents. To solve the problem, we propose a novel PRF method in which initial document retrieval is performed to search shorter documents (passages) about 30 seconds, which are simply divided according to the number of utterances, and related terms are extracted from them. We show that the proposed PRF outperforms conventional PRF on longer document retrieval tasks, and then, show the effect of combination of retrieval results obtained by the original query and expanded query by the proposed PRF.

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。これまでの音声ド

キュメント検索の主な研究対象は TREC SDR に代表されるニュース [1][2] や NTCIR-9[3] および NTCIR-10[4] に代表される講演であった。これらにおいて、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][5] や、クエリ拡張 [2][5] などが研究されている。

クエリ拡張とは、ユーザが入力した初期クエリに対していくつかの関連語を加えることでより検索に適したクエリを生成する手法である。代表的なクエリ拡張手法として、Web ページ [6] やシソーラスを用いる手法 [7]、および初期

¹ 龍谷大学理工学研究科
Graduate School of Science and Technology, Ryukoku University

² 龍谷大学理工学部
Faculty of Science and Technology, Ryukoku, University

a) nishio@nlp.i.ryukoku.ac.jp

b) nanjo@nlp.i.ryukoku.ac.jp

c) yoshimi@nlp.i.ryukoku.ac.jp

クエリによる検索結果から関連する語を抽出する擬似適合性フィードバック [7] などがあげられる。本研究では、ある程度の長さを持つ講演や講義の音声（以下、講演音声ドキュメント）を対象とした擬似適合性フィードバックを研究する。

擬似適合性フィードバックとは初期クエリによる音声ドキュメントの検索結果の上位を擬似的に適合ドキュメントとみなして関連語を抽出する手法である。この手法をよく機能させるためには、適合ドキュメントとみなすドキュメントが適切であることが重要である。具体的には、初期検索結果の検索性能が低すぎないこと、および、ドキュメントができるだけ関係する話題のみで構成されていること、の2点が重要である。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、講演音声ドキュメント検索を考えた場合は、検索タスクとして、講演や講義そのものを検索する講演検索タスクと、その一部分を検索対象とするパッセージ検索タスクがある。このうち、講演検索タスクやパッセージ検索のうちでも比較的長い音声区間を対象とする検索タスクでは、ある程度の検索精度が得られることがわかっている [8][9]。すなわち擬似適合性フィードバックに用いる検索結果の上位のドキュメントに、真の適合ドキュメントがある程度含まれている。しかし、検索で得られる適合ドキュメントは長く、複数の話題が含まれている。クエリに強く合致する真の適合箇所はドキュメントの一部であることがほとんどである。このため、このようなドキュメントからの関連語の抽出は難しい。テキストドキュメントを対象とした検索では、章、節、段落区切りが明確であるため、これらの情報を元に真の適合箇所を特定し、そこから強く関連する語を抽出することが考えられる。しかし、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、このようなアプローチをそのまま適用することができない。

これに対して、本論文では、講演音声をあらかじめ30秒程度の単位に機械的に区切って短めの音声ドキュメント群を用意しておき、それを対象として初期検索を行って関連語抽出を行う手法を提案する。提案手法は、長い音声ドキュメント検索、例えば講演検索や比較的長い音声区間を対象とするパッセージ検索において、検索したい単位とは異なる、適当な長さの単位に自動分割された発話単位を話題のまとまりとみなして、そこから関連語抽出を行うものであり、この点において新しい。さらに、対象とする音声ドキュメント以外の知識源、例えばWEBやシソーラスなど、を必要としない単純な手法であるといえる。

本論文では、提案する擬似適合性フィードバックによるクエリ拡張の効果を明らかにした上で、拡張したクエリによる検索結果と拡張前クエリによる検索結果を組み合わせることの効果を示す。

本論文の構成について述べる。2章では本研究で用いた

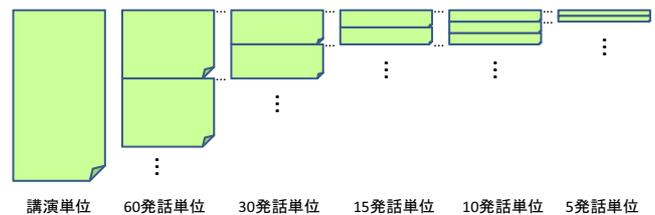


図1 各発話単位の関係

検索評価用テストコレクションについて述べる。3章では検索システムについて述べる。4章では擬似適合性フィードバックの手法および効果について述べる。5章では講演音声ドキュメント検索のための擬似適合性フィードバックの手法および効果について述べる。6章では拡張前クエリと拡張後クエリの併用手法とその有効性について述べる。7章で結論を述べる。

2. 講演音声ドキュメント検索

2.1 検索評価用テストコレクション

本研究では、NTCIR-9 SpokenDoc で用いられたテストコレクション [3] を用いて研究を行う。これは、日本語話し言葉コーパス [10] (以後、CSJ と記す) の学会講演 987 件と模擬講演 1715 件の合計 2702 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。CSJの各講演音声に対する音声認識結果が必要であるが、これについてはCSJの2702件の講演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 [8] に含まれているもの(認識率は65%から95%)を利用した。本研究ではパッセージ検索の対象として、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 [8] でも利用されていた60発話単位、30発話単位、15発話単位の他に、新たに10発話単位、5発話単位を採用する。60発話単位は講演の先頭から順に60発話ごとに区切ったものであり、この各区間が検索対象のドキュメントとなる。30発話単位と15発話単位、10発話単位、5発話単位についても60発話単位と同様に講演の先頭から順にそれぞれ30発話、15発話、10発話、5発話ごとに区切ったものである。各発話単位の関係を図1に示す。

クエリは自然言語文で記述された125件のテキスト(NTCIR-9 SpokenDocのdry run用クエリ39件およびformal run用クエリ86件)である。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おり、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。具体的には各クエリに対する正解ドキュメントとして、この適合ラベル(R)がつけられた区間を一部でも含む自動分

割された発話区間とした。

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) [11] を用いる。これは式 (1) で求められる。AP(\mathbf{q}_k) は南條ら [12] と同様の方法で求めた。

$$11\text{ptAP} = \frac{1}{N} \sum_{k=1}^N AP(\mathbf{q}_k) \quad (1)$$

3. 検索システム

本論文ではベクトル空間モデル [11] に基づくドキュメント検索システムを用いて、擬似適合性フィードバックの効果を検証する。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。本論文では、ベクトル間の類似度に SMART[13] を用いる。これはあるクエリ \mathbf{q} とドキュメント $D_i (1 \leq i \leq N)$ の類似度を、 \mathbf{q} と D_i のそれぞれの索引語 $t_k (1 \leq k \leq m)$ の正規化出現頻度 q_{t_k} および d_{i,t_k} を用いて、式 (2) で与えるものである。

$$\text{SMART}(\mathbf{q}, D_i) = \sum_{k=1}^m (q_{t_k} \cdot d_{i,t_k}) \quad (2)$$

ただし、

$$d_{i,t_k} = \begin{cases} \frac{1 + \log(\text{tf}_{i,t_k})}{1 + \log(\text{avtf})} & \text{if } \text{tf}_{i,t_k} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$q_{t_k} = \begin{cases} \frac{1 + \log(\text{qtf}_{t_k})}{1 + \log(\text{avqtf})} \log \frac{N}{n_{t_k}} & \text{if } \text{qtf}_{t_k} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

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、 \mathbf{q} 中での t_k の出現数、 avqtf は \mathbf{q} に含まれる単語の出現数の平均を表す。N は検索対象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数を表す。

本研究では、先行研究 [9] に従って、索引語を動詞と名詞の基本形とし、検索エンジンには汎用連想計算エンジン GETA*1を用いてドキュメント検索システムを設計する。

*1 汎用連想検索エンジン GETA (<http://geta.ex.nii.ac.jp>)

クエリ \mathbf{q} が与えられたとき、全てのドキュメント D_i について \mathbf{q} との類似度 $\text{SMART}(\mathbf{q}, D_i)$ 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し、評価を行う。

4. 擬似適合性フィードバック

4.1 クエリ拡張の概観

本論文では情報検索の方法としてベクトル空間に基づく方法を用いる。ベクトル空間モデルに基づく検索システムではクエリは単語の出現頻度を要素としたベクトルで表される。初期クエリのベクトルを \mathbf{q} 、クエリに付け加える新しい索引語のベクトルを \mathbf{q}_a 、重みを α とした場合、拡張されたクエリのベクトル \mathbf{q}_n は式 (5) で表される。ベクトル空間モデルではこの \mathbf{q} を \mathbf{q}_n にすることがクエリ拡張といえる。

$$\mathbf{q}_n = \mathbf{q} + \alpha \mathbf{q}_a \quad (5)$$

一般的に $\alpha < 1$ であるが、本実験では \mathbf{q}_n の要素を整数とするため、式 (6) に基づき β を正の整数としてクエリ拡張を行う。すなわち初期クエリ \mathbf{q} に対し再重み付けを行いクエリ拡張をする。

$$\mathbf{q}_n = \beta \mathbf{q} + \mathbf{q}_a \quad (6)$$

4.1.1 適合性フィードバック

適合性フィードバックとは、得られた検索結果のうち、どのドキュメントが検索意図に適合し、どのドキュメントが適合でないかをユーザが検索システムに入力することにより、クエリベクトル \mathbf{q} を \mathbf{q}_n に修正するものである [11] (式 (7))。

$$\mathbf{q}_n = \omega_0 \mathbf{q} + \omega_1 \mathbf{d}_r - \omega_2 \mathbf{d}_n \quad (7)$$

このとき、 \mathbf{d}_r と \mathbf{d}_n はそれぞれ適合ドキュメントの集合と不適合ドキュメントの集合に含まれる単語の出現頻度を各要素としたクエリベクトルとしたものである。また、 $\omega_0, \omega_1, \omega_2$ は 0 以上の定数である。

4.1.2 擬似適合性フィードバック

ユーザにより適合ドキュメントであるかどうかの判断を必要とせず、ユーザとのインタラクションなしに関連語を抽出し、クエリ拡張を行う手法を擬似適合性フィードバック (PRF: Pseudo Relevance Feedback) という [11]。本研究では、この PRF を音声ドキュメント検索に用いてクエリ拡張を行う。この様子を図 2 に示す。PRF では、はじめにクエリ \mathbf{q} を用いて検索結果を得る。次に得られた検索結果の上位いくつかを擬似的に適合ドキュメント集合とし、これらのドキュメントから関連語を抽出し、初期クエリに追加することでクエリ拡張を行う。このとき、不適合ドキュメントは用いない。従って、PRF は以下の式 (8) で表される。

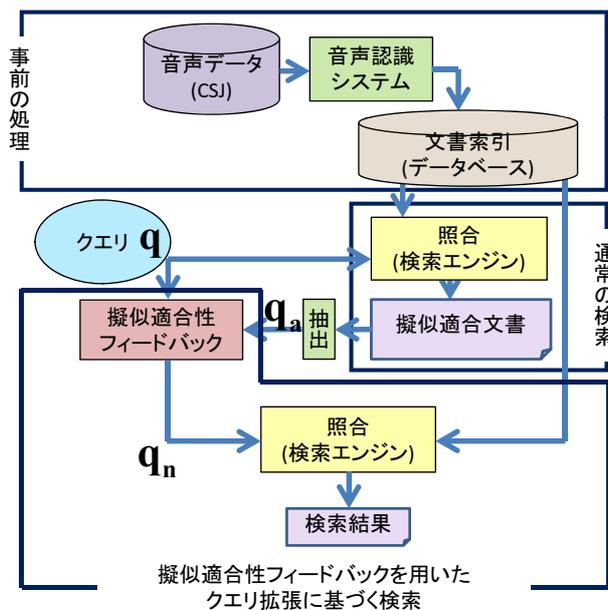


図 2 PRF を用いた音声ドキュメント検索の概観

$$\begin{aligned}
 \mathbf{q}_n &= \omega_0 \mathbf{q} + \omega_1 \mathbf{d}_r \\
 &\approx \frac{\omega_0}{\omega_1} \mathbf{q} + \mathbf{d}_r \\
 &\approx \beta \mathbf{q} + \mathbf{q}_a
 \end{aligned}
 \tag{8}$$

なお、本実験では、関連語の抽出を単語の出現頻度に基づいて行う。具体的には初期クエリから得られた擬似的な適合ドキュメント集合をひとつのクエリとみなして、式(4)に基づいて各語の出現頻度 q_{tk} を求め、この値が高いものから順に抽出する。

4.2 擬似適合性フィードバックの実験条件

PRF には、以下のパラメータがある。

- 擬似適合ドキュメント数
- 関連語数
- 重み

擬似適合ドキュメント数は、初期クエリ \mathbf{q} で得られた検索結果の上位何件を擬似的に適合ドキュメントとみなすかについてのパラメータである。関連語数は、適合ドキュメントから式(4)を用いて関連語抽出を行う際に何語抽出するかについてのパラメータである。重みは式(6)における初期クエリに対する重み β である。本研究では、125件のクエリを用いて Leave-one-out の交差検定を行った。具体的には、124件のクエリを用いて11点平均精度の平均が最大となるように各パラメータの推定を行い、残りの1件の検索実験にその推定パラメータを用いた。その際、擬似適合ドキュメント数は1件から5件までの1件刻み、関連語は10語から50語まで10語刻み、 β は1から10まで1刻みで変化させて実験を行った。これを全ての分割セットに対して行った。なお、本論文では関連語抽出を行う際に、ドキュメント出現頻度が1 ($n_{tk} = 1$) の語を除外した。こ

表 1 実験結果

検索単位	Baseline	PRF
講演単位	0.531	0.470
60 発話単位	0.327	0.305
30 発話単位	0.260	0.271
15 発話単位	0.187	0.205
10 発話単位	0.149	0.175
5 発話単位	0.104	0.123

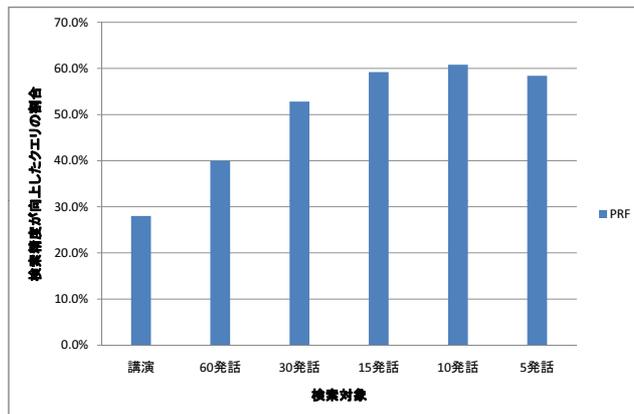


図 3 検索精度が向上したクエリの割合

れは、このような語を追加しても、すでに得られているドキュメント以外を求められないためである。

4.3 擬似適合性フィードバックの効果

まず、初期クエリでの各発話単位の検索結果について述べる。初期クエリでの検索結果の11点平均精度の全クエリでの平均値を求めた。検索結果を表1 (Baseline 列) に示す。講演単位、60 発話単位、30 発話単位、15 発話単位、10 発話単位、5 発話単位での11点平均精度の平均値はそれぞれ0.531, 0.327, 0.260, 0.187, 0.149, 0.104であった。これらは、検索対象ドキュメントが短いほどその検索精度が低いことを示している。

PRF による拡張後クエリでの各発話単位の検索結果を評価した。結果を表1 (PRF 列) に示す。

30 発話単位、15 発話単位、10 発話単位、5 発話単位の検索では11点平均精度の平均値はそれぞれ0.260から0.271, 0.187から0.205, 0.149から0.175, 0.104から0.123と向上したものの、講演単位、60 発話単位では11点平均精度の平均値はそれぞれ0.531から0.407, 0.327から0.305と低下した。これらは検索対象となるドキュメントが短いほどPRFの効果が大きく、ドキュメントが長いほどPRFは効果が小さいまたは悪影響が大きいことを示している。

初期クエリでの検索結果と比べてPRFによる拡張後クエリでの検索結果で11点平均精度がどれだけのクエリで向上したか調べた結果を図3に示す。ドキュメントが長い講演単位、60 発話単位では、PRFによって検索精度が低下しているクエリのほうが多いことが分かる。

表 2 PRF の影響の例

検索単位	講演単位	15 発話単位
クエリ	アボリジニーはどこに国にいるか。	
関連語	日焼け 生る 粋 鮫 時差 島国 紫外線 ビーチ サンタ 真夏	アボリジニー 大陸 移住 生る 粋 原住民 陸上 先週 旗 迫害
11ptAP	1.000 → 0.861	0.394 → 0.688

PRF の影響の例（講演単位で検索精度が低下し、15 発話単位で検索精度が向上したクエリ）を表 2 に示す。講演単位では関連語中に当該クエリの検索意図を含む大きな話題から選ばれた語が多いことが分かる。これに対し、15 発話単位では関連語中に当該クエリの検索意図に関連する語が数語あることがわかる。これは、講演単位での検索はうまく行われているが、検索結果として得られたドキュメントが、ある話題を中心に複数の小話題を含むものであり、PRF の際に当該クエリと関連しない小話題の語が関連語として抽出されているためと考えられる。15 発話単位での検索は、検索結果として得られたドキュメントが小さく複数の話題を含みにくいため、PRF の際に当該クエリに大きく関連する語が抽出されたためと考えられる。

5. 講演音声ドキュメント検索のための擬似適合性フィードバック

講演音声ドキュメント検索の中で講演検索タスクやパッセージ検索のうちでも比較的長い音声区間を対象とする検索タスクでは、ある程度の検索精度が得られることがわかった。しかし、このような検索タスクで PRF を行った場合、検索結果で得られる適合ドキュメントは長く、複数の話題を含むため、クエリに強く合致する一部分から関連語を抽出することは難しい。テキストドキュメントを対象とした検索では、章、節、段落区切りの情報を元に真の適合箇所を特定し、そこから強く関連する語を抽出することができるが、音声ドキュメントには、タイトルや段落区切りなどは明示されないため、このようなアプローチをそのまま適用することができない。

これに対して、本論文では、講演音声をあらかじめ 30 秒程度の単位に機械的に区切り、分割された発話単位を話題のまとまりとみなして、そこから関連語抽出を行う手法を提案する。提案手法の概観を図 4 に示す。これを本論文では講演音声ドキュメント検索のための擬似適合性フィードバック (PRFL: Pseudo Relevance Feedback for Lectures) とよぶことにする。

5.1 実験条件

PRFL には、PRF と同様のパラメータの他に、関連語の抽出を行うための索引にどの発話単位を用いるかについてのパラメータがある。本研究では、関連語抽出を行うための索引に 15 発話単位、10 発話単位、5 発話単位の索引

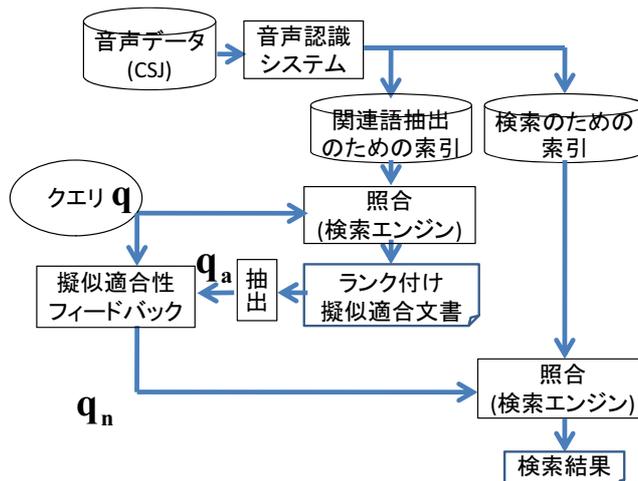


図 4 PRFL の概観

を採用した。このとき、それぞれのドキュメント長の平均は 35 秒、24 秒、12 秒となる*2。擬似適合ドキュメント数、関連語数、重みのパラメータは 4 章の実験で得られたものを用いた。

本研究では講演検索タスクやパッセージ検索のうちでも比較的長い音声区間を対象とする検索タスクだけではなく、短い音声区間を対象とする検索タスクでも実験を行った。

5.2 講演音声ドキュメント検索のための擬似適合性フィードバック

関連語抽出用索引として 15 発話単位、10 発話単位、5 発話単位を用いて PRFL を行った。検索結果を表 3 (PRFL (15), PRFL (10), PRFL (5)) に示す。PRFL (15), PRFL (10), PRFL (5) はそれぞれ、関連語抽出用索引に 15, 10, 5 発話単位を用いた結果を表している。PRFL (15) では、通常の PRF と比較して、講演単位、60 発話単位、5 発話単位での 11 点平均精度の平均値がそれぞれ 0.470 から 0.482, 0.305 から 0.325, 0.123 から 0.124 に向上したものの、30 発話単位、10 発話単位ではそれぞれ 0.271 から 0.261, 0.175 から 0.166 に低下した。PRFL (10) では、通常の PRF と比較して、講演単位、60 発話単位、15 発話単位、5 発話単位での 11 点平均精度の平均値がそれぞれ 0.470 から 0.489, 0.305 から 0.328, 0.205 から 0.210, 0.123 から 0.129 に向上した。30 発話単位では 0.271 から 0.270 と 11 点平均精度が低下しているものの、大きな差はなかった。PRFL (5) では、通常の PRF と比較して検索精度の向上がみられなかった。これらの結果は比較的長い音声区間を対象とする検索タスクで関連語抽出用索引として 15 発話単位や 10 発話単位が有効であることを示している。なお、スライドを利用する講演ではスライド 1 枚につき 1 つの話題について話されていると考えられ、その時間がおおよそ 30 秒から 1 分程度であることが多く、15 発話単位や 10 発話単位はこ

*2 単純に全講演の長さの合計を各ドキュメント数で割った平均

表 3 実験結果

検索単位	Baseline	PRF	PRFL (15)	PRFL (10)	PRFL (5)
講演単位	0.531	0.470	0.482	0.489	0.447
60 発話単位	0.327	0.305	0.325	0.328	0.301
30 発話単位	0.260	0.271	0.261	0.270	0.248
15 発話単位	0.187	0.205	(0.205)	0.210	0.195
10 発話単位	0.149	0.175	0.166	(0.175)	0.158
5 発話単位	0.104	0.123	0.124	0.129	(0.123)

* () は通常の PRF と同じ

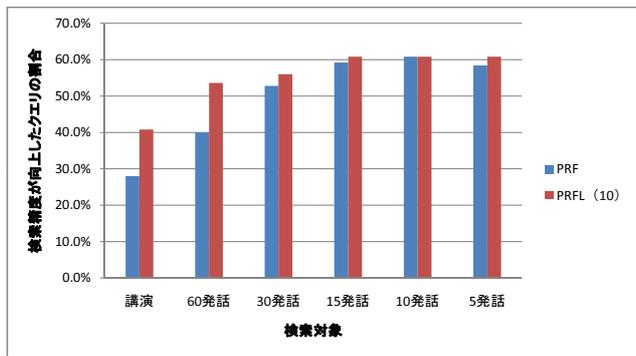


図 5 検索精度が向上したクエリの割合

れを擬似的にモデル化していると考えられる。また 5 発話単位は短すぎて初期検索の精度が十分でなく、うまく機能しなかったと考えられる。なお、本論文では長さを固定して一様に自動分割した発話区間を初期検索（関連語抽出）に用いている。分割長を可変にすることや話題の転換点の推定に基づいての分割などは今後の課題である。

初期クエリでの検索結果と比べて PRFL (10) による拡張後クエリでの検索結果で 11 点平均精度がどれだけのクエリで向上したか調べた結果を図 5 に示す。通常の PRF を行うよりも、すべての検索単位で 11 点平均精度が向上したクエリ数が増加していることがわかる。特に講演音声ドキュメントの検索で検索精度が向上したクエリの増加率が大きい。これらは、講演音声ドキュメント検索をする際には PRFL を行うことが特に有効であることを示している。また、パッセージ検索に対しても特にパッセージが長いときに PRFL を行うことが有効であることがわかる。

通常の PRF と比較して、PRFL では 11 点平均精度が向上しているものの、講演単位での検索では初期クエリよりも検索精度が低い。そのため、今後の課題として、長いドキュメントの検索に対してより頑健な PRF の手法の検討を行うことが挙げられる。

6. 拡張前クエリと拡張後クエリの併用

PRF や PRFL によるクエリ拡張では、擬似適合ドキュメントから関連語を抽出する。クエリによってはこの擬似適合ドキュメントに真の適合ドキュメントが含まれず、検索精度の低下の原因となる可能性がある。4.3 節では PRF による拡張後クエリでの検索結果で 11 点平均精度が低下

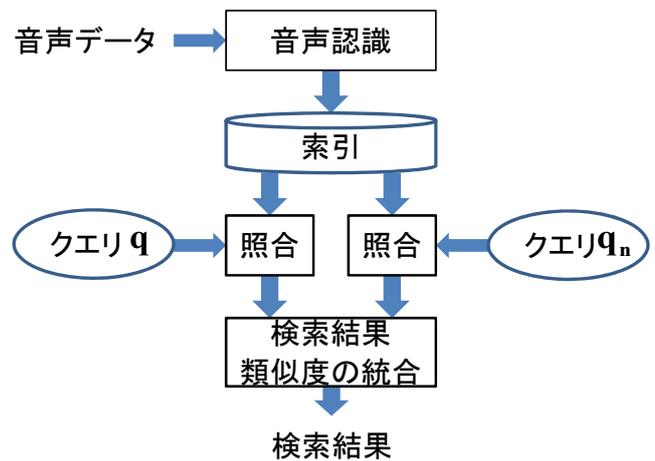


図 6 拡張前後のクエリ併用による音声ドキュメント検索の概観

したクエリがおおよそ 70% から 40% あることがわかった。5.2 節では PRFL による拡張後クエリでの検索結果で 11 点平均精度が低下したクエリがおおよそ 55% から 40% あることがわかった。

この問題に対処するために本研究では、初期クエリ（拡張前クエリ）での検索結果と拡張後クエリでの検索結果とを併用する [14][15][16]。これはクエリ拡張による精度低下を拡張前クエリでの検索結果で補うことを目的としている。すなわち、拡張前クエリ q とドキュメント D_i 間の類似度 $SMART(q, D_i)$ と拡張後クエリ q_n とドキュメント D_i 間の類似度 $SAMRT(q_n, D_i)$ とを組み合わせて検索を行う。本研究では、類似度の統合方法に線形対数補間を用いる。具体的には式 (9) に基づいて類似度を統合し統合後の類似度 $SIM(q, D_i)$ を得る。 λ は統合重みを表す。

$$SIM(q, D_i) = \lambda \cdot \log SMART(q, D_i) + (1 - \lambda) \cdot \log SMART(q_n, D_i) \quad (9)$$

こうして統合された類似度 $SIM(q, D_i)$ を用いてドキュメント D_i を類似度の降順に並べ替え最終的な検索結果とする。拡張前後のクエリの併用による検索の概観を図 6 に示す。

6.1 実験条件

拡張前クエリと拡張後クエリの併用では重み付け統合パ

表 4 実験結果

検索単位	Baseline	PRF	PRFL (10)	Base+PRF	Base-PRFL (10)
講演単位	0.531	0.470	0.489	0.542	0.542
60 発話単位	0.327	0.305	0.328	0.337	0.353
30 発話単位	0.260	0.271	0.270	0.276	0.280
15 発話単位	0.187	0.205	0.210	0.212	0.219
10 発話単位	0.149	0.175	(0.175)	0.174	(0.174)
5 発話単位	0.104	0.123	0.129	0.123	0.131

* () は通常の PRF と同じ

ラメータが必要である。具体的には、拡張前クエリ q とドキュメント D_i 間の類似度と拡張後クエリ q_n とドキュメント D_i 間の類似度の統合重み (式 (9) の λ) を 0.1 から 0.9 まで変化させ、125 件のクエリを用いて Leave-one-out の交差検定を行った。

6.2 拡張前クエリと拡張後クエリの併用の効果

拡張前クエリと PRF による拡張後クエリの併用の結果を表 4 (Base+PRF 列)、拡張前クエリと PRFL (10) による拡張後クエリの併用の結果を表 4 (Base+PRFL (10) 列) にそれぞれ示す。

拡張前クエリと PRF による拡張後クエリの併用 (Base+PRF) により、拡張前クエリ (Base)、拡張後クエリ (PRF) のいずれかのみを用いたときよりも高い、もしくは同等の 11 点平均精度が得られた。拡張前クエリと PRFL (10) による拡張後クエリの併用でも同様の傾向が得られた。併用においては PRF で拡張したクエリよりも PRFL (10) で拡張したクエリを用いる効果が大きかった。Base+PRFL (10) により、講演検索タスクで 11 点平均精度が 0.531 から 0.542、60 発話検索タスクで 11 点平均精度が 0.327 から 0.353 となった。

拡張前クエリでの検索結果 (Base) と比べて拡張前後のクエリの併用 (Base+PRFL (10)) での検索結果で 11 点平均精度がどれだけのクエリで向上したか調べた結果を図 7 に示す。比較のため、PRF の結果と、Base+PRF の結果も示してある。

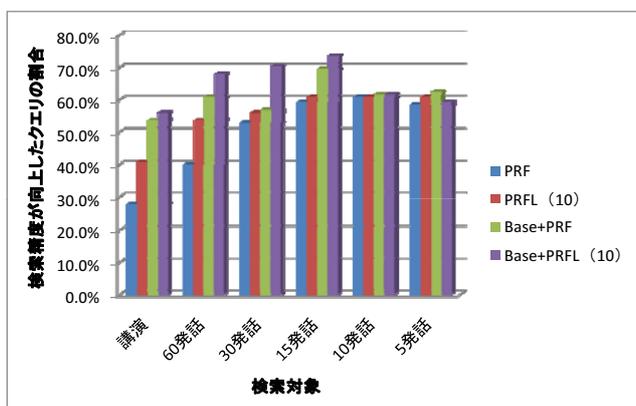


図 7 検索精度が向上したクエリの割合

拡張前後のクエリの併用によって検索精度が向上したクエリの割合も増加していることがわかる。特に講演検索タスクやパッセージ検索のうちでも比較的長い音声区間を対象とする検索タスクで大きく増加していることがわかる。

これらのことは、拡張前後のクエリの併用は有効であり、特に PRFL (10) による拡張クエリと拡張前クエリの検索結果の統合に効果がみられることを示している。PRFL を行った上で、その結果とベースラインの検索結果を統合することで、従来の PRF および検索結果の統合 (Base+PRF) よりも高い精度で検索でき、かつ、多くのクエリで改善が得られることが分かった。

7. 結論

CSJ の講演を対象とした検索タスクにおいて、PRF の効果を明らかにした。特に長い音声ドキュメントの検索タスクにおいて検索単位とは独立した適当な長さの発話単位を検索して得られた関連語を PRF に用いること (PRFL) の効果を示した。拡張前クエリと PRFL による拡張後のクエリを併用することの有効性を示した。

謝辞

本研究は科研費 (24500225) の助成を受けて行われたものである。

参考文献

- [1] Johnson, S. E., Jourlin, P., Jones, K. S. and Woodland, P. C.: Spoken Document Retrieval for TREC-8 at Cambridge University, *TREC* (1999).
- [2] Johnson, S. E., Jourlin, P., Jones, K. S. and Woodland, P. C.: Spoken Document Retrieval for TREC-9 at Cambridge University, *TREC* (2000).
- [3] Akiba, T., Nishizaki, H., Aikawa, K., Kawahara, T. and Matsui, T.: Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop, *NTCIR-9* (2011).
- [4] NTCIR-10: <http://research.nii.ac.jp/ntcir/ntcir-10/index-ja.html>.
- [5] 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).
- [6] Hasegawa, K., Sekiya, H., Takehara, M., Niinomi, T., Tsuge, S. and Hayamizu, S.: Toward improvement of SDR accuracy using LDA and query expansion for SpokenDoc, *NTCIR-9* (2011).
- [7] Jourlin, P., Johnson, S. E., Spärck, K. and Woodland, P. C.: General Query Expansion Techniques For Spoken

- Document Retrieval, *Proc. ESCA Workshop on Extracting Information from Spoken Audio* (1999).
- [8] 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, *IPSS-Journal*, Vol. 50, No. 2, pp. 501–513 (2009).
- [9] 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).
- [10] 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).
- [11] 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).
- [12] 南條浩輝, 弥永裕介, 吉見毅彦: 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, *IPSS-journal*, Vol. 53, No. 6, pp. 1654–1662 (2012).
- [13] 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).
- [14] Sheldon, D., Shokouhi, M., Szummer, M. and Craswell, N.: LambdaMerge: merging the results of query reformulations., *WSDM'11*, pp. 795–804 (2011).
- [15] Aslam, J. A. and Montague, M.: Models for metasearch, *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '01, pp. 276–284 (2001).
- [16] Montague, M. and Aslam, J. A.: Metasearch consistency, *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '01, pp. 386–387 (2001).