Vol.2018-SLP-125 No.8
2018/12/10

音声中の検索語検出における Web 検索と Word Vector を用いたリスコアリング方式

丹治遥[†1]　小嶋和徳[†1]　李時旭[†2]　南條浩輝[†3]　伊藤慶明[†1]

概要：音声データから特定のキーワードを検索する音声中の検索語検出（STD：Spoken Term Detection）の研究が盛んに行われている．検索精度を向上させるために，高順位候補を含むドキュメント内の全ての候補の距離を有利にする方式など[1][2]が提案されている．本稿では，同一講演内で話題の内容に関連してクエリと共起する単語をクエリの関連語と呼び，関連語は当該講演内に複数回出現すると仮定する．クエリの関連語を特定するため，本稿では Word2vec を用いた単語の分散表現が有効と考え，音声データの単語認識結果中の各単語を Word Vector 化し，クエリの Word Vector と比較し，類似度を求めることでクエリの関連語を取得する．一方，未知語クエリは単語認識結果に出現しないため Word Vector を算出できない．本稿では，クエリで Web 検索し得られたテキスト中の出現単語も Word2vec に用いてクエリの意味的情報を補い，未知語クエリの Word Vector を算出できるようにする．これにより，未知語クエリに対応させることができ，クエリの関連語を的確に求められると考える．以上のようにして，Web 検索と Word Vector を用いてクエリの関連語を特定し，関連語を含むドキュメント内の全ての候補の距離を有利にすることで検索精度の向上を図る．講演音声を対象とした NTCIR-10,12 の Formal Run の 2 種のテストセットを用いて評価した結果，両テストセットで検索精度が平均 3.2pt 向上した．先行手法と併用することで更に精度が平均 1.4pt 向上し，提案手法の有効性を確認できた．

キーワード：音声中の検索語検出，Web 検索，Word Vector，リスコアリング

Rescoring by Using Web Search and Word Vector for Spoken Term Detection

HARUKA TANJI[†1]　KAZUNORI KOJIMA[†1]　SHI-WOOK LEE[†2]
HIROAKI NANJO[†3]　YOSHIAKI ITOH[†1]

Abstract: This paper proposes a rescoring method for Spoken Term Detection (STD) using web search and word vector. The experimental results demonstrated the proposed method works well for open test collections that were distributed from National Institute of Informatrics (NII) for STD evaluation.

Keywords: Spoken Term Detection, Web Search, Word Vector, Rescoring

1. はじめに

近年，家庭での Blu-ray Disc レコーダや Web 上の動画投稿サイト等の普及に伴い，音声を含む大量のビデオデータを保存する機会が増加しており，それらのデータ中からユーザが所望する特定の区間を検索する機能に対するニーズが高まっている．この機能の実現のため，ビデオデータ中の音声情報を用いて検索語(クエリ)を検索する音声中の検索語検出(STD：Spoken Term Detection)の研究が盛んに行われている．国立情報学研究所が主催する NTCIR Workshop 9[3]が 2011 年，NTCIR Conferrence 10[4]が 2013 年，NTCIR Conference 11[5]が 2014 年, NTCIR Conference 12[6]が 2016 年に開催され, STD について様々な観点から評価された．

STD とは，音声ドキュメント内で一語以上からなるクエリが話されている位置を特定するタスクである．一般的な STD システムでは，大語彙連続音声認識システムを用いて検索対象の音声ドキュメントを予め認識し，その認識結果を用いて検索を行う．音声認識システムの単語辞書に登録されていない未知語がクエリとなると検索精度が低下する．この未知語のクエリに対応するため，単語より小さい単位のサブワードレベルでの認識結果を用いて照合を行う方式が一般的である[7]．クエリのサブワード系列と検索対象のサブワード系列を照合し，その照合距離の小さい順に候補として出力する．

先行研究[1]では，照合により得られた高順位候補は高い適合率を示すこと，及びクエリは特定のドキュメントに頻繁に出現することから，高順位候補を含むドキュメントにはクエリが複数含まれていると仮定し，高順位候補を含むドキュメント内の全ての候補の照合距離が小さくなるよう(有利になるよう)補正を行うことで，検索精度の向上を実現した．[2]では，高順位候補を含むドキュメントと内容が類似しているドキュメントにもクエリが複数含まれていると仮定し，類似ドキュメント内の全ての候補の照合距離が小さくなるよう補正を行うことで，検索精度の向上を実現した．

[8]では，STD において，クエリとよく共起する単語が検索

†1 岩手県立大学
Iwate Prefecctural University.
†2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
†3 京都大学学術情報メディアセンター
Academic Centar for Comptuting and Media Studies, Kyoto University

結果の候補の周辺に見つかれば，当該候補はクエリを含む正解である可能性が高いと考え，その候補の距離を有利にすることで検索精度の向上を実現した．この共起単語の情報はWeb 検索により得られるテキストから取得していたが，共起単語を正しく得られない場合はリスコアリングできず精度が向上しなかった．

音声ドキュメントは一般に話題, 対話, セッション, 講義, 講演単位等で分けられており, NTCIR の評価セットにおいても講演毎に分かれている．例えば，クエリを「岩手」とした場合，ある講演中に「岩手」と話されていれば，その講演では「岩手」に関する内容が話されていると想定できる.「岩手」に関連した「盛岡」「宮沢賢治」「わんこそば」などの単語も話される可能性がある．本稿ではこのように同一講演内で講演の内容に関連してクエリと共起する単語をクエリの関連語と呼び，クエリ及びクエリの関連語はクエリが出現する講演内に複数回出現すると仮定し，関連語を抽出した後，その関連語を含む講演内の全ての候補の距離を有利にすることで検索精度の向上を図る手法に取り組む．本研究では，クエリの関連語を見つけるために Word2vec[9][10]を用いる．Word2Vec は単語間の関連性をも表現できる単語の分散表現を求める手法であり，これにより求まった各単語の特徴ベクトル(以降, Word Vectror)を用いることで単語間の類似度を求めることができる．本稿では，音声ドキュメントを単語認識してその出現単語を Word Vector 化し，クエリと各単語の Word Vector を用いて類似度を計算し，クエリの関連語を求め，これを STD に用いる手法を提案する. このような研究はなされていない. このような関連語の抽出方式では，未知語クエリの問題がクリアできない．未知語クエリとは音声認識の辞書にないクエリのことであり，音声認識結果には含まれないため Word Vector を算出できないためである．本研究ではこの問題も扱う．具体的には Web 検索を併用する方式を採用する．Web 検索では検索単語に関するタイトルとスニペット（以降，Web テキスト）が複数出力される．この Web テキストには検索した単語が出現しており，その単語の意味や単語に関する話題などが含まれている．そこで，クエリ単語での検索結果の Web テキスト中の単語も Word Vector の学習に用いることで未知語クエリの単語的意味を学習し Word Vector を求めることができる．以上により，クエリと各単語の Word Vector を求め，クエリとの類似度を計算することで関連語が複数個得られる．

本稿では，以上のようにして選定した関連語を含む講演を抽出し，それらの講演内の全ての候補の照合距離を，その関連語の出現頻度の大きさに応じて有利になるよう補正(リスコアリング)する方式を提案し，その有効性を示す．関連語の抽出方法，補正の方法に新規性を有している．

本稿の構成は次の通りである．2.1 節では先行研究[1]の高順位候補を含むドキュメント優先方式について, 2.2 節では提案方式である Web 検索と Word Vector を用いたリスコアリング方式について述べる．3 章では提案方式の評価実験，先行研究[1]との比較統合について述べる．4 章で結論を述べる．

2. 提案方式

2.1 先行研究：高順位候補を含むドキュメント優先方式[1]

先行研究の高順位候補を含むドキュメント優先方式について概説する．はじめにで述べた通り高順位候補を含むドキュメントにはクエリが複数含まれていると仮定する．音声ドキュメントは講演$\Omega(A,B,C,\dots)$で構成されているとし，まず，STD を行った結果を講演毎に分類・順位付けを行う. 例えば，講演A内の高順位候補とクエリとの照合距離は小さい場合に，講演Aにはクエリを複数含んでいる可能性が高いと考える．そこで，高順位候補の照合距離を用いて講演A内の下位の候補区間の照合距離に対して, 以下の式(1)により調整(リスコアリング)を行う．$\alpha(0 \le \alpha \le 1)$は重み係数を表す。講演$A$の j 番目の発話がA内でk位であった場合の照合距離を$\mathrm{D}(\mathrm{A}_j,k)$とする。リスコアリング後の照合距離$\mathrm{D}'(\mathrm{A}_j,k)$は，その候補区間の元々の照合距離$\mathrm{D}(A_j,k)$と 1～$T$ 位$(1 \le \mathrm{T} \le k-1)$までの候補との照合距離$\mathrm{D}(\mathrm{A}_j,t)$の平均を線形結合することで求められる。

$$D'(\Omega_j,k) = \alpha D(\Omega_j,k) + (1-\alpha)\frac{1}{T}\sum_{t=1}^{T} D(\Omega_j,t) \qquad (1)$$

2.2 Web 検索と Word Vector を用いたリスコアリング方式

（1） Word2vec[9][10]

まず，本稿で用いた Word2vec について概説する．Word2vec[8][9]とは, ニューラルネットワークを用いた単語の特徴ベクトル化，すなわち単語の分散表現を求める手法である．この分散表現は単語の概念を表す低次元の密なベクトルで表される．学習テキスト中の各単語を周辺の単語から予測するタスク(疑似的な単語予測のタスク)を設定し，テキストデータを用いてニューラルネットワークで学習する．中間層における各単語の特徴を表す低次元ベクトルがその単語の重みであり，これを抽出することによって，単語の概念を表すベクトルを獲得する．周辺の単語の重みベクトルの和を中間層の値とする(周辺単語から中心単語を推定する)モデルを Continuous Bag-of-Words(CBoW)モデルと呼び，周辺の単語のうちの一つに対する重みベクトルを中間層の値とする(中心単語から周辺単語を推定する)モデルを Skip-gram モデルと呼ぶ．いずれのモデルも，入力層と中間層をつなぐ重み行列，つまり各単語に対する重みベクトルの集合が最終的に生成する単語分散表現(Word Vector)となる．これにより，単語を意味的空間上の一点に対応させることができ，単語に対する意味的な計算が可能となる．本稿では処理時間を考慮し，学習が Skip-gram よりも高速な CBoW モデルを用いた．

(2) 提案方式

次に，提案するリスコアリング方式について説明する．概要図を図 1 に示す．予め，検索対象の音声ドキュメントを音声認識システムを用いて単語認識する．クエリが与えられると，クエリと音声ドキュメントをサブワードレベルで照合を行い，その STD 結果を保持する．この検索結果に提案方式を適用する．以下，図の①~⑥の処理手順について説明する．

① Web テキストの取得

まず，クエリで Web 検索し，その検索結果の上位S件分の Web テキストを取得する．この Web テキストで Word2vec を学習すること(②)で，音声ドキュメントの単語認識テキストに出現しない未知語クエリの Word Vector を求めることができる．一方，Sを大きくしすぎた場合や検索結果のリンク先の本文ページの文章まで取得した場合，処理に時間を要するため，本稿では，Web 検索結果中のタイトルとスニペットのみとし，$S = 100$とした．

② Word2vec の学習

検索対象の講演音声の単語認識テキストと①で取得した Web テキストを Word2vec で学習し，両者に出現する単語の Word Vector を算出する．Web テキストと共に学習することで，クエリの意味的情報を補うことができ，より正確に単語の意味的な関連度合いを学習できると考える．

③ 関連語の選定

②によりクエリと各単語の Word Vector を求め，クエリとの類似度を計算することで関連語が複数個得られる．一方，音声ドキュメントには出現せず Web テキストのみに出現する単語とクエリとの類似度が想定外に高くなるケースが考えられ，その場合，関連語を適切に選定できないことが想定される．そのため，Word Vector を用いて求めた複数の関連語の中から，クエリの関連語として最も相応しい単語を選定する必要がある．本稿では，選定する関連語は名詞に限定する．そこで，クエリと類似度の高い名詞を関連語と決定するのではなく，類似度の高い複数の名詞単語を上位 N(=$100, 200, \ldots, 500$)個抽出し，クエリの関連語候補とする．[11]で示されたように，検索結果における最上位候補は最も適合率が高く，最上位候補を含む講演はクエリを含んでいる可能性が高いため，そのクエリの関連語は講演中に出現している可能性が高い．例えば，「音声認識」というクエリであれば，当該講演中で「音声認識」に関する内容が話されている可能性があり，「特徴量」「デコーダ」等の「音声認識」の話題に関連した単語も話されている可能性がある．これらの単語はクエリ周辺の発話を特徴付ける単語と考えられる．そこで，最上位候補を含む講演を対象とし，抽出したN個の関連語候補の中からクエリ周辺を特徴づける単語をクエリの関連語として選定する．具体的には，最上位候補を含む講演における複数の関連語候補に対しそれぞれの tf-idf 値を計算し，最も tf-idf 値の高い単語 1 個をクエリの関連語として選定する．複数個選定する場合も考えられるが，その検討は今後の課題とする．

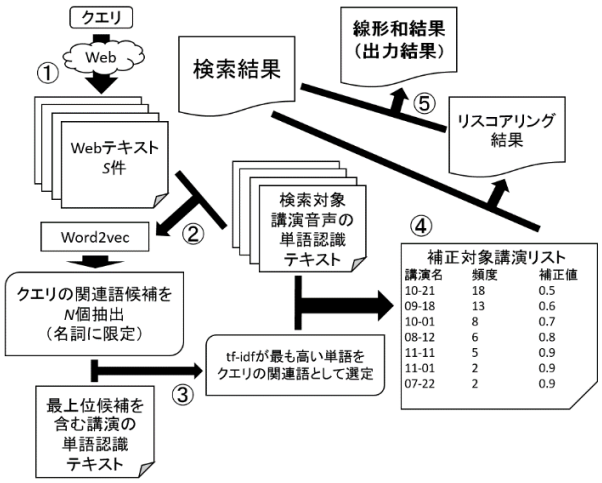

図 1 Web 検索と Word Vector を用いた
リスコアリング方式の概要図

④ 関連語を含む講演に対しリスコアリング

③で選定した関連語を含む講演はその関連語の出現頻度が高いほどクエリを含んでいる可能性が高いと考える．音声ドキュメントの単語認識結果に対し関連語で文字列検索することで，選定した関連語を含む講演を複数特定し，その特定した講演をリスコアリングの補正対象としリストに登録する．関連語の出現頻度が高い程リスコアリング時の補正効果が大きくなるよう補正値を設定する．リストに登録されている講演内の全ての候補に対して，以下の式(2)により，照合距離が小さくなるように補正を行う．$D(\Omega_j, k)$はリスト内の講演Ωの k 位の発話Ω_jの照合距離を表し，new$D(\Omega_j, k)$はリスコアリング後の照合距離を示す．$D(\Omega_j, k)$に補正値β($0.5 \leq \beta \leq 0.9$)を乗じて補正する。

$$\text{new}D(\Omega_j, k) = \beta \times D(\Omega_j, k) \quad (2)$$

この補正値の決め方は様々考えられるが，本稿では頻度順に 0.5，0.6，…，0.9 とし，頻度順位 5 番目以降の講演は全て 0.9 とした．

⑤ 線形和統合

④では，関連語を含む講演を抽出したが，クエリを含まない講演が抽出されるケースが考えられる．その場合は正しく補正されず，検索精度が低下する．そこで，このように間違って補正されるケースを考慮し，リスコアリング結果の照合距離に対し，元の検索結果の照合距離と線形和統合することで，適切な照合距離となるよう調整する．統合は以下の式(3)を用いて行う．$\gamma(0 \leq \gamma \leq 1)$は統合時の重み係数を表す．統合後の照合距離$newD'(\Omega_j, k)$は，④でリスコアリングした後の照合距離$newD(\Omega_j, k)$と元の照合距離D$(\Omega_j, k)$を線形結合することで求める．

$$\text{new}D'(\Omega_j, k) = \gamma \times newD(\Omega_j, k) + (1-\gamma) \times D(\Omega_j, k) \quad (3)$$

3. 評価実験

3.1 実験条件

音声ドキュメントの認識には DNN-HMM を用いて単語単位で認識を行った．音響モデルと言語モデルの学習データには，CSJ[12]の学会講演と模擬講演を合わせた 2,702 講演から評価に用いる 177 講演を除いた 2,525 講演のうち，偶数講演(1,255 講演，約 287 時間)を使用した．音響モデルは 3 状態の triphone で構成した．

DNN の学習に用いる音声特徴量は, 40 次元の FBANK とΔ, ΔΔの計 120 次元を用いた．音声特徴量の抽出条件は表 1 の通りである．DNN は Feedforward 型で，入力層は 1,320 ユニット，隠れ層は 2,048 ユニット×5 層，出力層は 3,238 ユニットで構築した．各層を RBM として Pre-training を行った後に RBM を連結して Fine-tuning を行うことで学習した．入力特徴量は FBANK 120 次元とし，中心フレームに前後 5 フレームを追加した 1,320 次元(11 フレーム×120 次元)とした．Kaldi[13]を用いて 3 状態の triphone を作成し，状態数は今回 3,238 状態となった．DNN の出力はこの triphone の 3238 個の状態の事後確率とした．

音声認識結果を状態系列とし，局所距離には状態間の音響距離[7]を用いた．クエリを triphone に変換し，同様に状態系列にした後，連続 DP(CDP : Continuous Dynamic Programming)照合を行うことで照合距離を求めた．

Web 検索エンジンは Google を使用した．Word Vector を用いた単語の特徴ベクトル化及び類似度算出には，Python 用トピックモデリングライブラリの gensim[14]で実装されている Word2vec を用いた．Word2vec の学習パラメータは，ベクトル次元数 : 200，文脈窓長 : 5，単語の最低出現頻度 : 1，学習係数 : 0.05 とした．

検索(STD)には，CPU : Intel Core i7-980EX，GPU : GeForce GTX 750 Ti，RAM : 12GB のマシンを使用した．

3.2 テストセット

評価には，表 2 に示す NTCIR-10，NTCIR-12 で用いられた Formal Run テストセットを使用した．NTCIR-10 では音声ドキュメントワークショップの講演音声(SDPWS : Corpus of Spoken Document Processing Workshop)の 104 講演(約 28.6 時間, 40,746 発話), NTCIR-12 では SDPWS の 98 講演(約 27.5 時間, 37,782 発話)が検索対象音声ドキュメントとして用いられた．クエリには, NTCIR-10 Formal Run で使用された 100 クエリ，NTCIR-12 Formal Run(Single term)で使用された 113 クエリを用いた．正解情報は，NTCIR オーガナイザから提供されたものを用いた．パラメータγとNについては，テストセット間での交差検証を行った．2.2 節の通り，パラメータβは $0.5 \leq \beta \leq 0.9$ の 0.1 刻みの値を取り，$S = 100$とする．

表 1：音声特徴量抽出条件

デジタル化	標本化周波数 16kHz 量子化 bit 数 16bit
特徴量	FBANK(40dim) + ΔFBANK(40dim) + ΔΔFBANK(40dim)
窓長	25 msec
フレームシフト	10 msec
窓関数	ハミング窓

表 2：テストセット

	NTCIR-10	NTCIR-12
検索対象データ	SDPWS104 講演 (約 28.6 時間 40,746 発話)	SDPWS98 講演 (約 27.5 時間 37,782 発話)
クエリ	Formal Run：100 種 (IV: 47, OOV: 53)	Formal Run：113 種 (IV: 72, OOV: 41)

3.3 評価指標

正解の判定は NTCIR 同様に発話単位で行い，クエリが発話内で一度以上話されていればその発話を正解とした．検索精度の評価には MAP(Mean Average Precision)を用いた．AP(Average Precision)は検索結果を上位から出力していき，正解が出力された時点での適合率を全正解で平均したものである．各クエリで AP を求め，それらを全クエリで平均したものが MAP となる．AP，MAP はそれぞれ以下の式(4)，(5)で求められる．クエリ$\boldsymbol{q}$に対する正解発話数を$\boldsymbol{C_q}$，$\boldsymbol{M}$は検索対象の総発話数，$\boldsymbol{\delta_i}$はバイナリ関数で，検索結果のi番目の発話が正解なら 1，不正解なら 0 となる．$precision(q,i)$はクエリqのi番目の検索結果出力時点での適合率である．Qはクエリ数を表す．

$$AP(q) = \frac{1}{C_q}\sum_{i=1}^{M}\delta_i \times precision(q,i) \quad (4)$$

$$MAP = \frac{1}{Q}\sum_{q=1}^{Q}AP(q) \quad (5)$$

3.4 提案方式の評価実験

パラメータγは 0.1 おきに，Nは 100~500 で変化させて実験を行った．結果を図 2 と図 3 に示す．図 2 は NTCIR-10，図 3 は NTCIR-12 のときの検索精度を示す．Baseline はリスコアリング方式適用前の結果を示す．それぞれのテストセットで以下のパラメータの組み合わせで最も検索精度が高くなった(Baseline からの向上値も示す).

NTCIR-10：$\gamma = 0.5, \mathrm{N} = 300$，3.3pt の向上(78.4%→81.7%)

NTCIR-12：$\gamma = 0.7, \mathrm{N} = 300$，4.1pt の向上(72.8%→76.9%)

どちらのテストセットでもNは 300，γは 0.5～0.7 に収束し，ほぼ同等のパラメータになったことから，本方式の頑健性を確認できた．

一方，NTCIR-12 では全てのγにおいて，安定して検索精度が向上したが，NTCIR-10 では$\gamma = 1.0$(リスコアリング後)のとき Baseline から約 1.0~2.0pt 減少した．適切に関連語を選定できず，クエリを含まない講演を誤って補正するケースが多かったことが原因と考える．

クエリ毎の結果を考察すると，計 213 クエリ中 Baseline で AP が 100%で本方式適用後も 100%のクエリを除いた 145 クエリのうち，74 クエリは検索精度が向上(内 OOV は 35 クエリ)，50 クエリは低下(内 OOV は 28 クエリ)，21 クエリは変化がなかった(内 OOV は 13 クエリ)．AP が向上したクエリで，例えば「アーティキュレーション」(OOV)の AP は 21.4pt 向上(67.7%→89.1%)した．選定された関連語は「発音」で，「アーティキュレーション」を含む講演で「発音」が複数出現しており，更に Web テキスト中でその意味を解説する記事が多かった．このため，これらの文章から単語的意味を学習できたと考える．AP が低下したクエリで，例えば「API」というクエリの AP は 16.6pt 減少(36.1%→19.5%)し，選定された関連語は「仕様」であった．Web テキスト中では API の仕様について書かれている記事が多かったが，「API」を含む講演には「仕様」が全く出現していなかったため，正しく学習できなかったと考える．検索精度が向上，低下するクエリの共通点や法則性は現段階では確認できず今後の課題とする．検索精度に変化がなかったクエリの中で「キタチャンキタロボ」(OOV)は，Web テキスト中に出現せず，Word2vec で学習できなかったため本方式が適用できなかった．(1 クエリのみ)

3.5 先行研究との比較・併用実験

提案方式(2.2 節)と先行方式(2.1 節)，及びそれらを併用した方式との比較を行った．その結果を図 4 に示す．併用における+は適用順を示す．MAP は NTCIR-10 と NTCIR-12 のテストセット間で交差検証(パラメータ設定)により求めた．

Baseline と各方式単体の検索精度を比較すると，以下のように MAP が向上した(左から先行方式，提案方式，括弧内は Baseline との比較を示す)．

NTCIR-10：1.3pt(78.4%→79.7%)，2.6pt(78.4%→81.0%)
NTCIR-12：3.9pt(72.8%→76.7%)，3.7pt(72.8%→76.5pt)
平均　　：2.7pt(75.4%→78.1%)，3.2pt(75.4%→78.6%)

提案方式は先行方式より平均 0.5pt 高い検索精度の向上を実現した．

Baseline と比べ両テストセットの平均で，先行+提案で 3.6pt，提案+先行で 4.6pt の向上となり，提案方式を適用後に先行方式を適用する場合が最も精度が高くなり，この場合，先行，提案の単体の良い方の精度を上回った．これは，提案方式を先に適用することで，選定した関連語を含む講演内の全ての候補が有利な照合距離となり，先行方式による補正で更に有利な照合距離となったためと考える．

先行+提案において最も検索精度が高くなったパラメータは，NTCIR-10 で$\gamma = 0.1, N = 400$，NTCIR-12 で$\gamma = 0.5, N =$

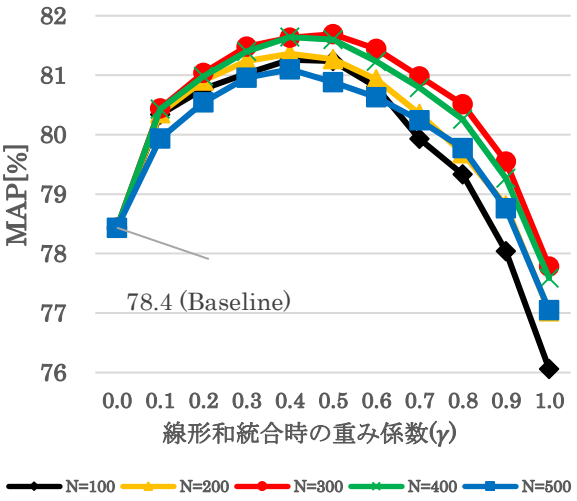

図 2：NTCIR-10 に提案方式を単体で適用した結果

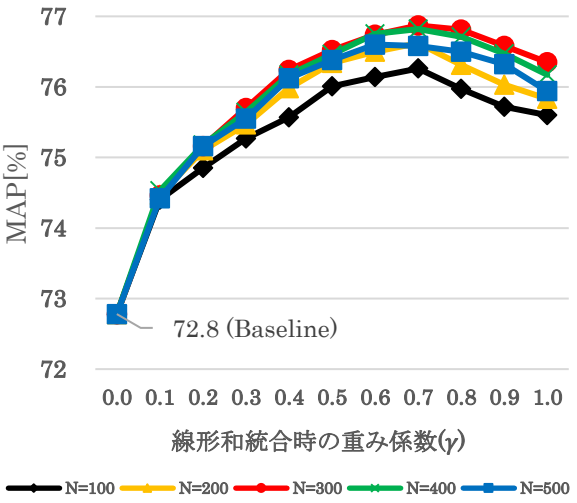

図 3：NTCIR-12 に提案方式を単体で適用した結果

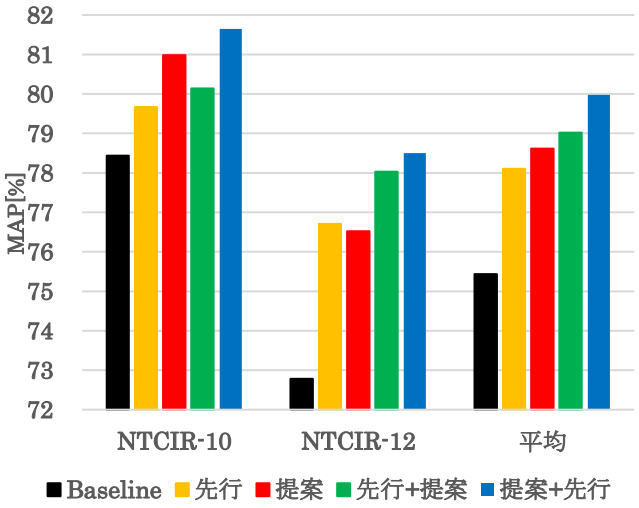

図 4：各方式の検索精度

300で，パラメータγに差があった．3.4 節で述べた通り，NTCIR-10 では適切に関連語を選定できず，クエリを含まない講演を誤って補正するケースが多かったため，併用による大きな補正効果が得られず，γは低い値に収束したと考えるが，詳細については今後の課題とする．

3.6 提案方式の処理時間計測

本方式は処理手順が多く，検索に時間を要すると想定される．提案方式において検索結果が得られるまでの時間を計測した(STD の検索時間は除く)．その結果を図 5 に示す．図に示す処理時間は NTCIR-10 と NTCIR-12 の計 213 クエリの(1 クエリ当たりの)平均の処理時間である．MAP はγ = 1.0のときの NTCIR-10 と NTCIR-12 の平均を示す．

Nを 100 増やす毎に平均で 0.73 秒増加している．最も検索精度の高いN = 300のときで 2.84 秒，その内の 2.18 秒は関連語を選定するための tf-idf の計算時間であった．この tf-idf の処理をせずに関連語を選定すれば更に処理速度が速くなるが，その具体的な手法の検討と検索精度への影響については今後の課題とする．

4. 結論

本稿では STD において Web 検索と Word Vector を用いてクエリの関連語を選定し，その関連語を含む講演内の全ての候補の距離を有利にするリスコアリング方式を提案した．検索精度においては，提案方式で 3.2pt の向上が得られ，提案手法の有効性を確認した．更に提案方式を適用後に先行方式を適用することで 1.4pt，トータル 4.6pt の向上が得られ，提案方式と先行方式を併用することの有効性も確認できた．提案方式のパラメータの関連語候補数Nは 300，線形和統合の重み係数γは 0.5~0.7 に収束し，本方式の頑健性が確認できた．一方，先行研究と併用時の NTCIR-10 でのパラメータの差については調査が必要と考える．今後は，Web 検索結果の取得件数Sを大きくし Word2vec の学習テキスト量を増やしたときの検索精度への影響，N = 400以上にすると精度が低下する原因の調査，リスコアリング時の補正値βの自動決定方法，選定する関連語の適切な個数について検討していく．

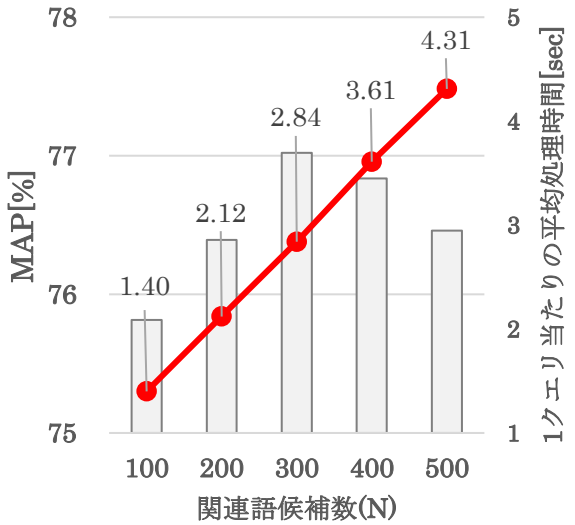

図 5：提案方式の処理時間

謝辞 本研究の一部は JSPS 科研費 18K11358 の助成を受けたものです．

参考文献

[1] 小嶋和徳，紺野和磨，田中和世，李時旭，伊藤慶明：音声中の検索語検出における同文書内の高順位候補を利用したリランキング方式，電子情報通信学会 D Vo1.J100-D No.1, pp70-80, 2017.
[2] 清水嘉乃，李時旭，小嶋和徳，伊藤慶明：音声中の検索語検出におけるドキュメント間類似度を利用したリスコアリング方式，情報処理学会第 80 回全国大会，5Q-08，pp.2-393--394，2018-3.
[3] T. Akiba, H. Nishizaki, K.Aizawa, T. Kawahara and T. Matsui : Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop, NTCIR-9 Workshop Meeting, pp.223-235,2011.
[4] T. Akiba, H. nishizaki, K. Aikawa, X. Hu, Y. Itoh, T. Kawahara, S. Nakagawa, H. Nanjo and Y. Yamashita : Overview of the NTCIR-10 SpokenDoc-2 Task, NTCIR-10 Workshop Meeting, pp. 573-587, 2013.
[5] T. Akiba, H. Nishizaki, H. Nanjo and G.I.F. Jones : Overview of NTCIR-11 Spoken&Doc Task, NTCIR-11, pp. 350-364, 2014.
[6] T. Akiba, H. Nishizaki, H. Nanjo and G.J.F. Jones : Overview of NTCIR-12 Spoken&Doc Task, NTCIR-12, pp. 167-179, 2016.
[7] 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方式における新しいサブワードモデルとサブワード音響間距離の有効性の検証，情報処理学会論文誌，vol48，no.5，pp.1990-2000，2007.
[8] 小田原一成，山下洋一：音声中の検索語検出における単語共起情報の利用，情報処理学会研究報告，2016-SLP-110，pp.1-6，2016.
[9] T. Mikolov，I. Sutskever，K. Ghen，G. Corrado，J.Dean : Efficient Estimation of Words and Phrases and their Compositionally, Advances in Neaural Information Processing Systems 26，pp.3111-3119，2013.
[10] T. Mikolov，K. Ghen，G. Corrado，J.Dean : Efficient Estimation of Word Representaions in Vector Space，Processing of the International Conference on Learning Representaions (ICLR)，pp.1-12，2013.
[11] 丹治遥，小嶋和徳，李時旭，南條浩輝，伊藤慶明：音声中の検索語検出における最上位候補を含む講演及びその類似講演優先方式，日本音響学会春季研究発表会，2-Q-17，pp.185-186，2018-3.
[12] National Institute for Japanese Language and Linguistics：Corpus of Spontaneous Japanese，http://pj.ninjal.ac.jp/corpus_center/csj/
[13] D. Povey，A. Ghoshal，G.Boulianne，L. Burget，O.N. Goel，M. Hannemann，P. Motlicek，Y. Oian，P. Schwarz，J. Silovsky，G. Stemmer and K.Vesely : The Kaldi Speech Recognition Toolkit，ASRU，2011.
[14] gensim topic modeling for humans : https://radimrehurek.com/gensim/index.html