

# 複数の入力説明文を用いた音声ドキュメントからの用語検索

森田 直樹<sup>†</sup> 南條 浩輝<sup>‡</sup> 馬 青<sup>†</sup>

<sup>†</sup> 龍谷大学 理工学研究科

<sup>‡</sup> 京都大学 学術情報メディアセンター

t15m007@mail.ryukoku.ac.jp, nanjo@media.kyoto-u.ac.jp, qma@math.ryukoku.ac.jp

## 1 はじめに

動画検索のための音声ドキュメントからの用語検索について研究をしている。これまでに、主に辞書や Wikipedia 等の「見出しとその定義文」を自身に含む文書を検索対象とした用語検索の研究が多くなされている [1],[2],[3]。動画検索のためにはタグが重要であるが、動画検索タグとなる専門的な用語は辞書に載っていないことが多く、タグ語を思い出せないときは、見たい動画にアクセスできない。そのため、用語の説明文から用語を推測する用語検索が必要である。動画検索におけるタグには動画中の語を用いることが自然である。したがって、動画検索のための用語検索においては、動画中の音声認識結果（音声ドキュメント）からの用語検索技術が重要である。我々は、求める用語の説明文を検索クエリとして、検索クエリと意味的に類似している音声ドキュメントの一部（パッセージ）を選択し、そのパッセージに求める用語が含まれると仮定して、パッセージから関連語の抽出を行い用語候補とする手法を用いて用語検索を行っている [4]。

本稿では、用語検索システムの入力である説明テキストの効果的な利用法について検討を行う。これまで、入力された説明テキストが複数文からなる場合でも、これらを全てまとめて一つの説明文として扱ってきた。実際用語検索においては、ユーザは思いつかない用語に対して、様々な観点から説明を加えると考えられる。したがって、説明テキストの各文は表している側面がそれぞれ異なる可能性があり、全てをまとめて一つの説明文として扱うのは適切でない可能性がある。このような背景に基づき本稿では、説明テキストを分割してそれぞれで用語検索し、その用語候補を統合する手法を検討する。

## 2 データ

本研究では、文献 [4] で用いたテストデータと同様のものを用いる。

## 2.1 音声ドキュメント

検索対象には、日本語話し言葉コーパス (CSJ: Corpus of Spontaneous Japanese) [5] の学会講演 987 件と模擬講演 1715 件の合計 2702 件の講演であり、音声認識結果には文献 [6] で用いられたものと同一のものを利用する。

## 2.2 検索クエリ

テストデータは文献 [4] で用いられたものと同様の地名とカタカナ語それぞれ 25 個ずつである。各用語に対して、3 文からなる説明文が付与されている。説明文の例を以下に示す。

説明文の例

京都:

日本の関西の都市。清水寺や八坂神社といった寺や神社の名所多い。古都と呼ばれ歴史的価値のあるものが多い。

## 3 検索システム

検索システムには、我々が文献 [4] で提案した用語検索システム (図 1) を用いる。これは、文書選択システム、関連語選択システム、用語の抽出システムの 3 つから構成されている。それぞれについて以下に述べるが、詳細は文献 [4] を参照されたい。

### 3.1 文書を選択

検索クエリと意味的に類似度の高いパッセージを選択する。検索対象である音声ドキュメントには句読点や段落といった情報がない。そこで、本研究では無音（ポーズ）で区切られた音声を発話と定義し、その発話の内容によらず、発話の数で意味のあるまとまりとして考え、これをパッセージとする [7]。本実験では、

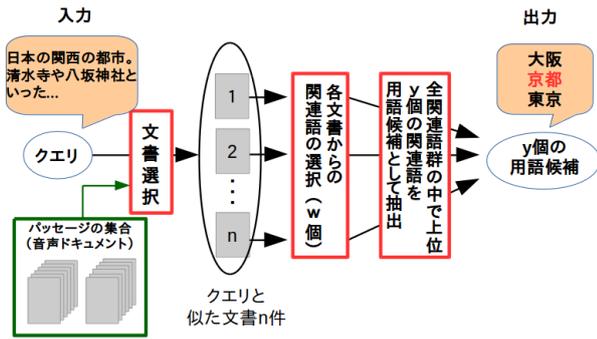


図 1: システムのイメージ図

文献 [4] で得られた結果を基に, 15 発話を 1 パッセージとして用いる. 文書選択システムはベクトル空間モデルに基づいて行う.

### 3.2 関連語の選択

選択されたパッセージから関連語を選択する. 関連語は, 文書選択で得られたパッセージ中から情報量等のスコアを基にして, そのパッセージと関連が深い語 (関連語) を選択する.

### 3.3 用語の抽出

用語の抽出は, 関連語群のスコアを与え, 用語らしさの高い順に候補として出力する. 検索クエリと関係がない語が文書選択の誤りや関連語選択の誤りにより, 関連語群に含まれる可能性があるため, ここで用語らしさの順に並び替える.

## 4 複数のクエリベクトルによる検索

先行研究 [4] では, 入力である検索クエリ (説明文) の 3 文を bag-of-words によるベクトルとして, まとめて 1 つの入力としてきた. しかし, 3 文のうち 1 つの文が長く, それ以外の文が短い場合には, 1 つのベクトルにまとめると長い文だけのベクトルとほぼ同様のものになってしまう問題があると考えられる. また, それぞれの説明文で用語候補としての順位が高くて, 1 つのベクトルにまとめてしまうことでクエリベクトルが平均化されてしまう問題もある. ユーザは様々な観点から説明を行い, 説明テキストの各文が表している側面がそれぞれ異なる可能性があり, これらをまとめて扱うのは適切でない可能性がある.

このような問題点から, 検索クエリの 1 つの説明に対して 1 ベクトルとし, 各用語候補のスコアを基に用語候補を出力する手法を提案する. 具体的には, 図 2 のように各文から 3 章と同様にして用語候補を抽出し, それらの用語候補のスコアを合計したものを, 新

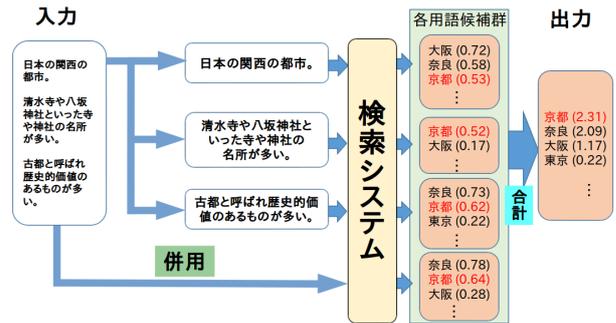


図 2: 複数のクエリベクトルによる用語検索

たなその用語候補のスコアとするというものである. これにより, 得られた各ベクトルで候補順位が高いものが最終的な用語候補として出力される. 今回の実験では, 3 文の説明文を用いるため, これを 3vector と表記する. また, 3 文の説明文だけでなく, これをまとめた説明テキスト (ベースラインの検索で用いられるもの) も用いる手法 (3vector+) も調査する. 図 2 は, 3vector+ の例である.

## 5 用語候補のスコアの正規化

各用語候補のスコアの正規化について述べる. 4 章では, 各用語候補のスコアをそのまま合計していた. しかし, 例えば図 3 のような場合には問題がある. この例では, 各文で用語候補を抽出した結果, 2 つ目と 3 つ目の文から得られた用語候補の順位に正解となる用語の「アメリカ」が 1 位となっている. しかし, 他の文や 3 文全体の結果と比べるとスコアが全体的に低く, このスコアのまま合計してしまうと, 1 位となっているものが半数であるにもかかわらず, これらのスコアが用語候補の順位にあまり影響を及ぼさない. この問題を解決するために, 式 (1) のようにして, 正規

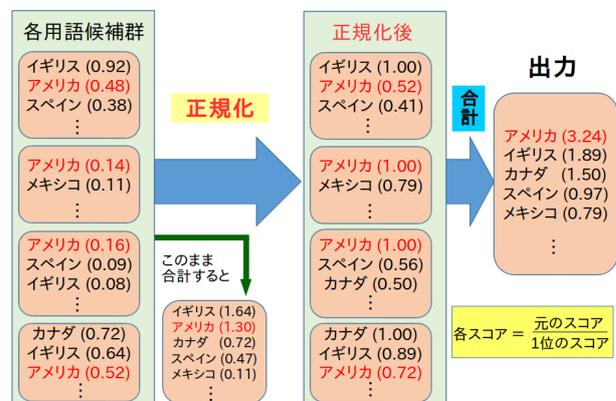


図 3: 用語候補の正規化

化を行う。

$$\hat{S}_{q_i, t_k} = \frac{S_{q_i, t_k}}{S_{q_i, t_1}} \quad (1)$$

具体的には、 $i$  番目の 1 つの文を入力ベクトルとして得られた用語候補のスコア  $S_{q_i, t_k}$  をそのベクトルで候補の順位が 1 位であった用語候補のスコア  $S_{q_i, t_1}$  でそれぞれ割ったものを  $\hat{S}_{q_i, t_k}$  とする。すなわち、1 位のスコアで各スコアを割ることで、各文から得られた用語候補のそれぞれの 1 位のスコアが最大値 1 となる。このような正規化を行い、得られたスコアを合計して用語候補を出力する。これにより、各入力文で得られた結果を平等に扱うことができ、問題が解決できると考えられる。

## 6 評価実験

### 6.1 評価指標

用語検索の評価は正解用語が出力された順に基づいて行う。本タスクは正解用語が 1 つであるため、検索において正解が 1 つだけの場合に適した評価指標である平均逆順位 (MRR: Mean Reciprocal Rank) を用いる。これは、式 (2) で定義されるものであり、正解用語が出力された順位の逆数の平均値である。値が大きいほど精度が高いことを示し、最大値は 1 である。

$$MRR = \frac{1}{Q_N} \sum_{q=1}^{Q_N} \frac{1}{tRank_q} \quad (2)$$

ここで  $tRank_q$  は検索クエリ  $q$  に対して、正解となる答えが用語候補として出力されたときの順位であり、 $Q_N$  は検索クエリの個数である。

本研究では平均逆順位を拡張した平均逆ページ順位 (MRPR: Mean Reciprocal Page Rank) [4] (式 (3)) も用いる。

$$MRPR = \frac{1}{Q_N} \sum_{q=1}^{Q_N} \frac{1}{tPageRank_q} \quad (3)$$

ここで  $tPageRank_q$  は検索クエリ  $q$  に対して、正解となる答えが用語候補として出力されたページである。

実際の学習支援の状況では、検索結果の用語候補を複数提示してユーザに求める回答となる用語を選択させる支援を想定している。この場合、求める用語は必ず候補の 1 番目ではなく、上位の複数に含まれていれば大きな問題ではないと考えられる。この指標では 1 ページに 10 個の用語候補が提示されると仮定し、何ページ目に正解となる用語が提示されたかを評価する。候補の 1 番目や 10 番目に提示された場合は MRPR は 1.0 となる。また 11 番目は 2 ページ目であるので 0.5 となり、21 番目は 0.333.. となる。

## 6.2 実験結果

表 1: 地名タスクの用語検索の結果

地名	baseline	3vector	3vector +
アメリカ	43	28	30
イギリス	9	1	1
イラン	151	*	389
エジプト	5	15	7
カナダ	139	366	188
シドニー	1	1	1
シンガポール	231	532	312
スペイン	1	6	1
ドイツ	17	7	12
メキシコ	73	323	132
モンゴル	*	*	*
ラスベガス	2	10	3
吉祥寺	41	58	51
群馬	17	21	19
広島	7	7	5
高崎	6	8	6
成田	11	6	7
静岡	459	124	158
千葉	9	7	7
中国	1	4	3
東京	33	14	15
熱海	44	146	88
八王子	813	35	57
名古屋	50	186	81
京都	237	63	65
再現率	96%	92%	96%
MRR	0.1840	0.1419	0.1920
MRPR	0.4770	0.4999	0.5110

正解となる用語の出現順位を表す

\* : 順位が 1000 位以内に正解がなかった

表 1, 表 2 に地名タスクとカタカナ語タスクの結果を示す。50 個の用語の説明文を用いて用語検索を行った。それぞれの説明文を入力として、正解となる用語が出力された順位を用いて評価を行う。これまでの研究では、入力ベクトルは 3 文を 1 つにまとめたものであり、その結果を baseline として、1 文ごとにベクトルとして用いた場合との比較を行う。表 1 の 3vector は、1 文ごとのベクトルを入力として実験を行った結果であり、3vector+ は、さらに 1 つにまとめたもののベクトルを追加して実験を行った結果である。

表 1, すなわち地名タスクにおいて baseline と比較すると、再現率と MRR が 3vector の場合に検索精度が低くなっているが、3vector+ では MRR, MRPR が向上していることが確認できる。地名タスクにおいて、入力ベクトルには 1 文ごとのベクトルだけでなく、1 つにまとめたベクトルを用いることで検索精度の向上が可能であることを確認できた。なお、1 文だけのベクトルの場合、クエリに含まれる単語が少ないために、うまくパッセージを選択できず、精度が悪くなると考える。表 2, すなわちカタカナ語タスクにおいては、クエリを分割する効果は確認できなかった(ただし大きな悪影響もなかった)。

表 3, 表 4 に、5 章で述べた正規化を用いた効果を示す。3vector の場合は、両タスクとも MRR が向上

表 2: カタカナ語タスクの用語検索の結果

カタカナ語	baseline	3vector	3vector +
アルゴリズム	10	13	5
アーティスト	*	643	*
カリキュラム	5	5	5
キーワード	3	6	2
コイル	167	294	197
コスト	201	865	239
コミュニケーション	8	10	9
コーパス	85	17	22
サリン	2	2	2
サンプル	181	175	142
スピーカー	198	461	254
ターゲット	189	*	439
ノード	141	891	400
バイオリン	12	25	17
ビット	*	232	436
プライド	357	*	*
プリンター	132	127	93
マラソン	3	3	3
マラリア	5	46	22
ユーザー	116	53	58
レントゲン	596	*	899
オリーブオイル	*	*	*
スターバックス	*	*	*
デシベル	*	*	*
パスポート	*	*	*
再現率	76%	72%	76%
MRR	0.0777	0.0617	0.0820
MRPR	0.3255	0.2787	0.3085

正解となる用語の出現順位を表す

\* : 順位が 1000 位以内に正解がなかった

されており、正規化に効果がみられることがわかる。3vector+において正規化を行うことで地名タスクでベースラインからの大きな向上を得た。カタカナ語タスクでは、大きな精度の向上は確認できなかったものの悪影響はなかった。これらのことより、本提案手法の有効性を確認できた。カタカナ語において、大きな効果は確認できなかった原因として、検索クエリが適切でなかった可能性(そもそも検索精度が低い)があると考えられる。

今後の課題として、候補の 1000 以内に正解がなかった用語への対応や、ユーザに質問の変更をうながす対応を行うような研究をしたい。また、出力された用語でも候補の順位が低いものも多く、用語の抽出に課題が大きいことが分かった。今後は、これらの改良を行っていきたい。

## 7 おわりに

複数のベクトルを入力とした音声ドキュメントからの用語検索の研究を行った。1 文ごとに入力ベクトルを生成し、それらに対する用語候補群のスコアを用いる手法、またスコアを正規化したものを用いて用語候補を得る手法を検討し、その有効性を確認した。

## 謝辞

本研究は科研費(課題番号 25330368, 15K00254)の助成を受けた。

表 3: 3vector における正規化の効果

地名	正規化なし	正規化あり
MRR	0.1419	0.1790
MRPR	0.4999	0.4655

カタカナ語	正規化なし	正規化あり
MRR	0.0617	0.1091
MRPR	0.2787	0.2907

表 4: 3vector+における正規化の効果

地名	正規化なし	正規化あり
MRR	0.1920	0.2195
MRPR	0.5110	0.5260

カタカナ語	正規化なし	正規化あり
MRR	0.0820	0.0831
MRPR	0.3085	0.3142

## 参考文献

- [1] 粟飯原俊介, 長尾真, 田中久美子.: “意味的逆引き辞書『真言』”, 言語処理学会第 19 回年次大会発表論文集, pp.406–409, 2013.
- [2] Qing Ma, Ibuki Tanigawa, and Masaki Murata.: “Retrieval Term Prediction Using Deep Learning Methods”, The 30th Pacific Asia Conference on Language, Information and Computation (Paclic 30), 2016.
- [3] 山島祥子, 大庭隆伸, 阪内澄宇.: “音声入力による人名想起支援:その言語的特徴の分析”, 音響学会講演論文集(春季), 3-4-7, 2014.
- [4] 森田直樹, 南條浩輝, 山本凌紀, 馬青.: “音声ドキュメントを検索対象とした用語検索”, 情報処理学会論文誌(2017年3月掲載予定).
- [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] Tomoyosi Akiba, Kiyooki 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”, IPSJ-Journal, vol.50, No2, pp.501–513, 2009.
- [7] 西尾友宏, 南條浩輝, 吉見毅彦.: “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”, 情報処理学会論文誌, Vol.55, No.5, pp.1573–1584, 2014.