

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 4 月 30 日現在

機関番号：13701

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500151

研究課題名（和文） 音声認識による話題語の継時的提示に関する研究

研究課題名（英文） Study on time varying display of keywords using speech recognition

研究代表者

速水 悟（HAYAMIZU SATORU）

岐阜大学・工学部・教授

研究者番号：90345794

研究成果の概要（和文）：連続的なメディアコンテンツに対して、音声認識結果として得られるテキスト情報をもとに、再生時点での要約情報を時間的に変化させる新たな手法を提案した。とくに、話題の変化と提示するキーワード選定の問題を同時に解決する手法として、統計的なトピック言語モデルの適用を提案し、映像に対して字幕とキーワードを付与し、その話題の変化を複数のキーワードによって示すシステムとして実現し、その有効性を明らかにした。

研究成果の概要（英文）：This study proposes a new method which shows time varying keyword information at the time of replaying continuous media content from speech recognition results. Using a statistical topic language model, captions with multiple keywords which shows topic changes are added to video. The method was evaluated by the implemented system and the effectiveness was shown.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	700,000	210,000	910,000
2011 年度	1,000,000	300,000	1,300,000
2012 年度	900,000	270,000	1,170,000
年度			
年度			
総計	2,600,000	780,000	3,380,000

研究分野：総合領域

科研費の分科・細目：情報学、知覚情報処理・知能ロボティクス

キーワード：音声情報処理

## 1. 研究開始当初の背景

## (1) 社会的背景

音声認識による映像メディアへの自動字幕付与は、聴覚障害者や高齢者への支援として期待されてきた。また、会議や講義などの記録を再生し内容を把握する際の支援として、音声認識技術の活用が期待されていた。

## (2) 学術的背景

関連する国内・国外の研究には、音声認識による発話のテキスト化とともに字幕として

の整形を行うものや、議会での発言をできるだけ忠実に文字に書き起こし、人手で修正するための議事録の作成支援システムなどがあつた。またテキスト処理による自動要約の研究も様々な手法が検討されていた。

## 2. 研究の目的

## (1) 研究目的の概要

本研究は、映像メディアの音声認識結果に対して、発話内容を短時間に把握できるような情報を抽出し、字幕とともに提示する技術の

開発を目的とした。

(2) 本研究の技術的な特色

音声認識結果から、時間とともにゆるやかに変化する話題や内容を表すキーワード群を抽出する手法を提案することと、キーワード群を字幕に併せて提示する方法の有効性を実験的に明らかにすることを特色とする。

3. 研究の方法

(1) 研究方法の概要

話題の変化を表すキーワード群による要約提示用のテキスト生成手法と、音声認識結果からの話題語抽出手法と、認知心理学的な評価実験による有効性の評価の3項目に重点を置いた。また経済とスポーツのような大きな話題の変化ではなく、全体の中で話題や内容が小さく連続的に変化するようなコンテンツへの適用をねらいとした。

(2) 評価実験の方法

コンテンツ内の話題推移の推定とキーワード群抽出の実験においては、領域の違いや話題展開形式が異なる新聞記事、講演、ニュース解説を対象とした。評価実験においては、映像に字幕とキーワード群を付与した閲覧システムを作成し、被験者実験を行った。キーワード群の抽出においては、複数のトピック言語モデルの結果を統合する方法を新たに提案した。また音声認識誤りの影響については Web 文書の利用の有無などの異なる実験条件でキーワード群を抽出して比較を行った。

4. 研究成果

(1) トピック境界とキーワード推定

話題の変化とそれを表すキーワード群の抽出手法として、統計的な言語モデルを利用する手法を提案した。

具体的には、トピック適応言語モデルを複数用いてトピック境界推定を行い、次いで得られたトピックごとに、そのトピックに対応するモデルを用いてキーワードを抽出する手法を提案した。

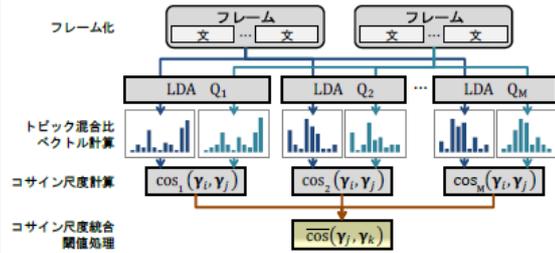


図2 複数トピックモデルの統合によるトピック境界推定

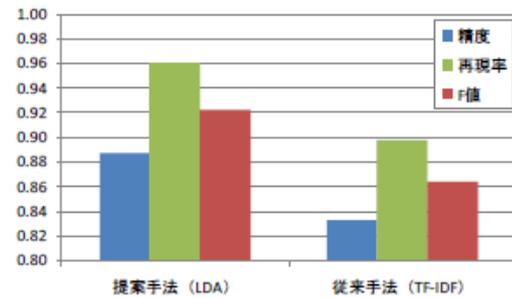


図3 キーワード抽出実験の結果

図1に、トピック境界推定とキーワード抽出を用いた字幕提示システムの構成図を示す。図2には、複数トピックモデルの統合によるトピック境界推定の概念図を示す。図3には、キーワード抽出実験の結果を示す。

本手法は文をまとめたフレームを単位として、LDA (Latent Dirichlet Allocation) と呼ばれるトピック適応言語モデルを用いて、各フレームのトピックの特徴を、トピック混合比ベクトルとして推定するものである。フレーム間の変化については、複数のトピックモデルの結果を統合することで、より頑健なトピック境界の推定を行う。

新聞記事を用いたトピック境界推定実験によって、提案手法の有効性を示した。モデ

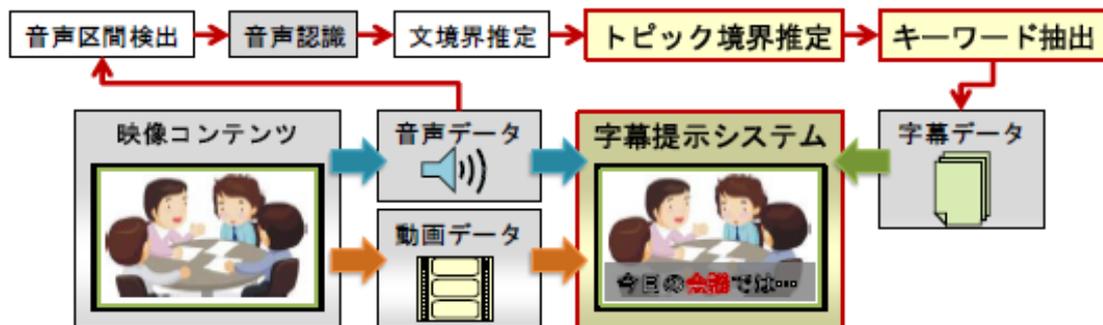


図1 トピック境界推定とキーワード抽出を用いた字幕提示システム

ルの学習には毎日新聞 2008 年 1 年分、評価には毎日新聞 2007 年の記事から、約 200 記事（約 2 万文）を無作為に抽出・並び替えを行ったデータを用いた。なお、学習データには記事ジャンルが付与されており、その総数は 16 ジャンルである。このデータに対し、トピック境界推定を行い、精度、再現率、F 値によって評価を行った。またキーワード抽出においても、従来手法である TF-IDF 値を用いるものと比較して、提案手法がより高い精度と再現率が得られることを示した。

## (2) トピック推移の解析

時間的にゆるやかな内容変化をとともなうコンテンツや、複数のトピックを同時に含むコンテンツを対象として、トピック適応言語モデルを用いてトピックの時間的な推移の解析を行った。

対象とする文書のフレーム化を行い、各フレームに対して LDA を用いてトピック混合比ベクトルを推定した。トピック混合比ベクトルによるトピック境界推定を行うとともに、トピック混合比の高いものから順に、潜在的なトピックの時間的な推移を解析した。

とくに頻出単語とキーワードという二通りの表現を用いて潜在トピックの意味づけを試みた。新聞記事およびニュース解説の書き起こしテキストを対象とした解析を行い、評価用データの各フレームに含まれる文や単語と、モデルの潜在トピックの頻出単語およびキーワードとを比較し、その関連性を調べるとともに、トピック混合比ベクトルの時間的な変化について考察した。

図 4 は、ニュース解説の書き起こしテキストにおけるトピック混合比の推移の例を示したものである。横軸が文書のフレーム番号、縦軸がトピック混合比である。

各フレームの内容と推定されたトピック混合比の推移を検討した結果、おおむね適合することがわかった。また、頻出単語は一般名詞を多く抽出できるため字幕内での強調に適しており、キーワードは固有名詞を多く抽出できるため動画の具体的内容を連想するのに適していることがわかった。

(3) 音声認識信頼度の推定とキーワード抽出  
音声認識誤りへの対策として、平均相互情報量の推定に Web 文書を利用する手法を提案した。

音声ドキュメントの認識信頼度を推定する手法として、2 単語間の自己相互情報量 (PMI: Pointwise mutual information) を用いる方法が提案されており、共起情報を基にした単語の文脈一貫性の計算に用いられている。これに対して、PMI の共起の計算に音声ドキュメントの話題に応じた Web 検索の結果を用い、話題に特化した学習を逐次行うことで認識信頼度の向上を図った。また、PMI をもとに各単語のスコアを算出するキーワード抽出手法を提案した。

PMI の共起の計算に、音声ドキュメント毎の Web 検索結果を用いた。これにより、話題に合う共起関係を多く取得することができ、大規模コーパスを用いて共起の計算を行う場合に比べ、各音声ドキュメントに適した認識信頼度の推定が可能となる。

評価セットの音声ドキュメントとして、模擬対話 2 件(それぞれ約 2 分のデータ)と、CSJ 2702 講演から 10 講演のデータの一部、合計 12 件のデータを使用した。認識信頼度の推定は、評価セットの各音声ドキュメントに対して、(1) 共起情報を毎日新聞記事データ集(2007 年度)から取得し、PMI を平滑化したもの、(2)(3) 評価データ毎に共起情

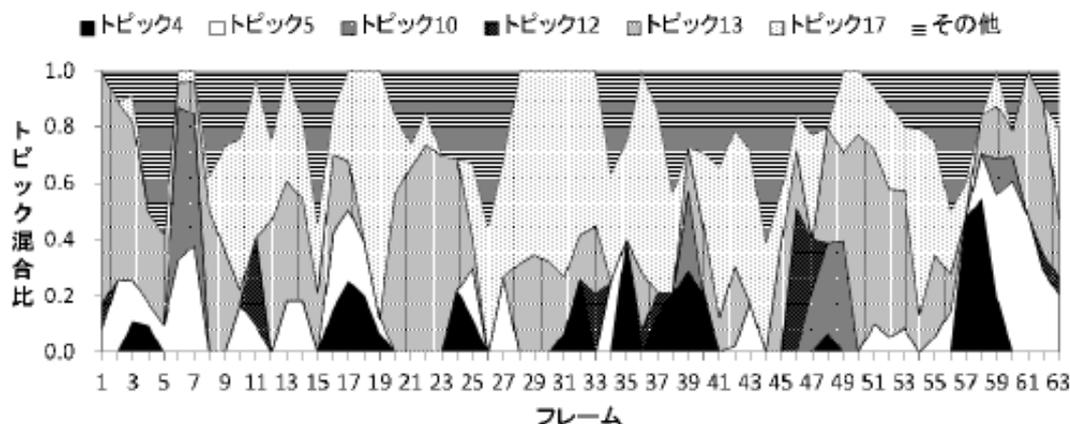


図4 ニュース解説におけるトピック混合比の推移

表1 認識信頼度と単語認識率の誤差

(N, M)	(1) 毎日新聞	(2) Web	(3) Web_weight
(5, 2)	0.260	0.198	0.084
(10, 5)	0.111	0.108	0.038
(20, 5)	0.072	0.046	0.032
平均	0.148	0.117	0.051

報を Web から取得し平滑化 PMI を適用 (Web 及び Web\_weight), の 3 条件で認識信頼度を算出し, 各ドキュメントのフレーム毎の単語認識率との二乗誤差を比較した.

評価セットの音声ドキュメントとして, 模擬対話 2 件(それぞれ約 2 分のデータ)と, CSJ 2702 講演から 10 講演のデータの一部, 合計 12 件のデータを使用した. 音声ドキュメント認識信頼度の計算におけるフレーム分割のフレーム長 N 及びフレームシフト M は, (N, M) = (5, 2), (10, 5), (20, 5) の 3 パターンとした. 内容語として, 名詞のみを用いた.

表 1 に, 認識信頼度と単語認識率との誤差の関係を示す. 表 1 より, どのフレーム分割においても, 毎日新聞での共起学習に比べ, Web 情報を用いた場合の認識信頼度と単語認識率の二乗誤差が小さくなった. これは, Web 情報を共起学習に用いることで, 音声ドキュメントに関連した単語を多く取得するとともに, 関連しない認識誤り単語の共起学習を抑えられたためであると考えられる. 特に Web\_weight で小さい誤差が示された.

また, キーワード抽出の実験として, 評価セットの各音声ドキュメントにおいて, TF-IDF によるキーワードと, 毎日新聞, Web, Web\_weight の共起情報を用いた PMI 加算値から得られたキーワードの計 4 結果を比較する被験者実験を行った. 被験者数は 20 名である. キーワード群は抽出結果上位 10 単語とし, 4 結果のキーワード群の適切さを比較し, 順位付けをさせた. 12 件の評価セットの平均順位は, TF-IDF が 2.67 であるのに対し, Web\_weight が 1.71 と TF-IDF を上回り, 良好な結果が得られた. さらに実際の映像に対して字幕とキーワード群を付与し, その話題の変化を提示するシステムを作成し, 被験者実験によって, 提案した手法の有効性を示した.

#### 5. 主な発表論文等

(研究代表者, 研究分担者及び連携研究者には下線)

#### [雑誌論文](計1件)

Masanao Okamoto, Kiichi Hasegawa, Sho Sobue, Akira Nakamura, Satoshi Tamura, Satoru Hayamizu, Topic-based generation of keywords and caption for video content, 査読有, Proceedings of the Second APSIPA Annual Summit and Conference, pp.605-608, Singapore (2010-12).

[http://www.apsipa.org/proceedings\\_2010/pdf/APSIPA121.pdf](http://www.apsipa.org/proceedings_2010/pdf/APSIPA121.pdf)

#### [学会発表](計4件)

川瀬徹也, 田村哲嗣, 速水悟, 音素誤り傾向を用いた音声認識誤り訂正の検討, 電子情報通信学会 2013 年総合大会, ISS-SP-377, p.219, 岐阜大学, 2013

原謙介, 関谷英樹, 田村哲嗣, 速水悟, Web 情報に基づく認識信頼度の推定とキーワード抽出, 日本音響学会 2013 年春季講演論文集, 3-P-27c, pp.235-236, 東京工科大学, 2013

新家太桜, 長谷川貴一, 竹原正矩, 関谷英樹, 田村哲嗣, 速水悟, トピック適応言語モデルを用いたトピック推移の解析, 第 6 回 音声ドキュメント処理ワークショップ, 豊橋技術科学大学, 2012

長谷川貴一, 岡本昌直, 祖父江翔, 中村明, 田村哲嗣, 速水悟, 複数トピックモデルを用いたキーワード抽出, 第 5 回 音声ドキュメント処理ワークショップ, 豊橋技術科学大学, 2011

#### [図書](計0件)

#### [産業財産権]

出願状況(計0件)

取得状況(計0件)

#### [その他]

#### 6. 研究組織

##### (1)研究代表者

速水 悟 (HAYAMIZU SATORU)

岐阜大学・工学部・教授

研究者番号: 90345794

##### (2)研究分担者

田村 哲嗣 (TAMURA SATOSHI)

岐阜大学・工学部・助教

研究者番号: 10402215

##### (3)連携研究者