

話題語に着目したニュース音声の要約法の検討*

堀 智織 岩崎 淳 古井 貞熙 (東工大)

1. はじめに

近年、大語彙連続音声認識の技術は著しく向上し、朗読調の音声であれば単語正解精度 90% 以上でほぼ実時間の認識が可能となった。現在では、放送ニュースや講演などの実世界の音声を認識対象とする検討が盛んになり、放送ニュースへの自動字幕付与や講演の自動議事録作成といった応用面に対する期待が高まっている。しかし、大語彙連続音声認識の結果をそのまま字幕や議事録にするのでは、冗長な情報が多く有効ではない。放送ニュースの字幕では、ニュースの進行や人間の読解速度に伴う時間的制約により画面表示できる文字数が制限されるため、文字数を圧縮する必要がある。また、議事録では発話内容の重要な部分を簡潔に表現することにより、自動検索などに用いるインデクシングや抄録としての価値を付与できる。これらは、音声認識の結果を要約処理することにより実現される。

これまで我々は、ニュース音声の認識結果から話題を表す語(話題語)を自動的に抽出することを検討し、種々の話題抽出尺度の比較検討を行ってきた[1]。本稿では、この話題語に着目して放送ニュース音声を要約する手法を提案する。本手法では、音声要約を「認識単語列から話題語とその間の単語を日本語としてもっともらしくつながるように抜き出す処理」として捉え、動的計画法を適用する。本手法によって生成した要約文を 8 人の被験者が評価した結果について考察する。さらに、作成した要約文の尤度を用いて、N-best 文候補をリスコアし、音声認識への寄与を調べる。

2. 音声要約のアプローチ

要約技術は自然言語の分野で盛んに研究されているが、その多くは対象となる文集合(段落)から、重要と思われる文を直接抜き出す手法である。接続詞や特有な表現を手掛かりに抄録から主題文を抽出したり[2]、見出しの中で使われている名詞を含む文を本文中から抜き出す方法[3]、単語の頻度により重要な文をいくつか取り出し、その文の語尾を体言止めにしたり、丁寧な言葉を言い切りの形に直したりして文を短くする方法などがある[4]。

本稿で提案する音声要約手法では、文集合から文を抽出するのではなく、各文に対して話題語とその間の単語をもっともらしくつながるように抜き出し要約文を生成する。ここでは、各単語の話題語らしさを表す話題語スコアと抜き出された単語列のもっともらしさを表すスコア(言語尤度)を基準に、スコア最大の部分単語列を動的計画法によって求める。従って、不要な単語を有効に取り除くことができ、かつ、情報の欠落が少ないことが期待できる。

話題語らしさを表す話題語スコアは、先の検討[1]で最も精度の高かった以下の尺度を用いる。

$$I(w_i) = g_i \log \frac{G_A}{G_i} \quad (1)$$

但し、

- w_i : ニュース音声中的名詞
- g_i : (注目している特定の) ニュース音声中的名詞 w_i の頻度
- G_i : 全データベース中の名詞 w_i の頻度
- G_A : 全データベース中の総名詞数 ($= \sum_i G_i$)

である。また、名詞以外の単語に対する話題語スコアは一

定値を与えるものとする。

言語尤度は単語 trigram で与えるものとして、 N 個の単語からなる認識単語列より抜き出された部分単語列を $V = v_1, v_2, \dots, v_M (M < N)$ とするとき、要約処理は次式を最大にする \hat{V} を求める問題となる。

$$S(V) = \sum_{m=1}^M \{\log P(v_m | v_{m-1} v_{m-2}) + \lambda I(v_m)\} \quad (2)$$

このような問題は動的計画法によって効率的に解くことができる。

ここで、 λ は話題語スコアと言語尤度間のバランスをとる係数であり、 λ が大きい時は話題語を数多く抽出する方に重点が置かれ、小さい時は単語のつながりの良さに重点が置かれる。

3. 要約スコアによるリスコアリング

誤認識が少ないほどもっともらしい要約文を抜き出すことができると仮定すれば、式(2)の要約スコアを音声認識部にフィードバックして認識精度の向上に寄与させることが考えられる。具体的には、音声認識によって N-best 文候補を求め、各候補を音響尤度、言語尤度、及び、要約したときの要約スコアを組み合わせるリスコアし、それが最大となる文候補とその要約文を出力する。

4. 評価実験

4.1 実験条件

提案手法を放送ニュース音声の認識によって評価する。評価データとして、語彙サイズ 20000 で未知語率 1.7% の放送ニュース 10 記事(計 48 文)を用いる。

4.2 音声認識システムの構成

【特徴抽出】

音声データを 12kHz, 16bit でデジタル化し、フレーム長 32ms, フレーム周期 8ms で対数パワーと 16 次元の LPC メルケプストラムおよびそれらの 1 次の回帰係数(計 34 次元)を抽出する。さらに発話毎のケプストラム平均正規化を行う。

【音響モデル】

Tree-based clustering による状態共有化を行った不特定話者音素文脈依存 HMM(2106 状態)を用いる。音声資料として ATR 音声データベース B セット、日本音響学会連続音声データベース、および同模擬対話データベースから、男性 53 名による 13270 発話を用いた。

【言語モデル】

単語 bigram, trigram を用いる。放送ニュース原稿テキスト 5 年分(1992 年 7 月から 1996 年 5 月)の約 50 万文を、形態素解析システム JTAG を用いて形態素に分解し、その形態素を単語として bigram と trigram を学習した。語彙サイズは 20000 語である。

【デコーダ】

単語グラフを中間表現とする 2 パスデコーダを用いる。第一パスでは HMM と bigram を用いてフレーム同期のビームサーチを行い、単語グラフを生成する。このとき、単語間の音素文脈依存も考慮する。また、音素グラフによる仮説制限[5]によって高速化を図っている。第二パスでは trigram を用いて単語グラフをリスコアする。

* A Study on Summarization of Broadcast News Speech Based on Topic Words

By Chiori Hori, Atsushi Iwasaki, and Sadaoki Furui (Tokyo Institute of Technology)

4.3 要約文の言語モデル

要約文の日本語としてのもっともらしさを評価する言語モデルとして N-gram を用いるが、これは認識時に使用する言語モデルとは別に用いる。

本来、要約文の N-gram の学習には、ニュース原稿から適切な基準を元に不要な単語を削除した大量の要約文を用いることが望ましい。しかしながら、現在そのような要約文で構成された大量のテキストは利用可能でないため、現段階では以下の手法で作成したテキストを用いる。

まず、8人の被験者により、ニュース原稿約50文中の各単語を下記の3段階で分類した。

- (1) 重要：情報として重要である
- (2) 不要：削除されても文意に影響が無い
- (3) 普通：どちらにも該当しない

この結果を元に、ニュース原稿から「不要」と選択された単語を機械的に取り除き、さらに文末を体言止めにする等の処理を施したテキストを疑似的な要約文として作成した。しかしながら、これらの処理は機械的に行われるため、日本語の表現として適切でないものが含まれてしまう。そこで、不適切な表現の影響を削減するため、疑似要約文に元のニュース原稿を1対1で混合して言語モデルを作成した。

4.4 要約文の評価

要約文は、第2パスで得られた音声認識結果を第2節で述べた手法を用いて作成した。その結果得られた要約文に対し、原文から要約文に抽出されるべき単語、および、原文に対する要約文の文意という2点に関して、8人の被験者によって評価した。但し、ここでは単語正解精度90%以上かつ25単語以上で構成される認識結果に対する要約文(要約率70%)を対象とした。

- (1) 抽出されるべき単語
評価用ニュース音声の書き起こしテキスト中の各単語に対して「重要」、「普通」、「不要」の3段階の分類を行い、「重要」と選択された単語が自動作成された要約文にどの程度含まれるかを評価した。その結果、抽出されるべき「重要」語の内86%が要約文に含まれることが確認できた。
- (2) 要約文の文意
要約文の文意について以下の3段階評価を行った。
 - i) 原文と同意である。
 - ii) 原文の文意に包含されている。
 - iii) 原文の文意とは異なる。

その結果「原文と同意」は30%、「原文を包含」は47%、「原文と異なる」は23%の割合となった。「原文と同意」ないし「原文を包含」で合わせて77%の割合で要約文に原文の文意を保持できたが、原文と異なる文意の要約文が23%存在することから、更に単語の意味情報などの上位の知識を利用することが必要と考えられる。

4.5 要約文例

要約結果の例として「原文と同意」、「原文に包含」、「原文と異なる」に高い割合で選択された文を原文と併せて以下に示す。

- (1) 原文と同意
【原文】公的介護保険制度をめぐるのは、保険の運営主体となる市町村の側が、厚生省の案では財政面などで不安が残るとして慎重な考えを示しており、法案の提出に向けては、市町村関係者の十分な理解を得られるかどうか最大の焦点となっています
【要約文】公的介護保険制度をめぐる保険運営主体となる市町村が厚生省で財政面などでとして慎重な

考えを示しており法案提出に向けて市町村関係者理解を得られるかどうか最大の焦点

- (2) 原文に包含
【原文】これを受けて与党側は住専処理法案などの今週中の採決を目指すことにしていますが、新進党は採決には抵抗する構えで、法案の衆議院通過をにらんだ攻防が続く見通しです
【要約文】これを受けて与党側は住専処理法案など採決を目指すことにしていますがあす採決で法案衆議院通過をにらんだ攻防が続く見通し
- (3) 原文と異なる

【原文】このうち加藤幹事長に対する参考人質疑では、加藤氏が、住専の融資先で倒産した鉄骨メーカー共和から、平成二年に一千万円を受け取ったと、加藤氏の元後援会長が発言している問題について、加藤氏自身がどのような答弁をするかが焦点となります
【要約文】うち加藤幹事長に対する参考人質疑で加藤氏が住専融資先で倒産した鉄骨メーカー共和から二千万円を受け取った加藤氏元後援会長が問題加藤氏自身が焦点となり

4.6 要約スコアによるリスクの効果

要約率を様々に変えて、第2パスで得られた N-best 文候補に対し要約スコアによるリスクを行った結果、baselineの単語誤り率が17.9%であるのに対し、要約率50%、70%、90%の要約スコアを適用した際の単語誤り率がそれぞれ17.9%、17.5%、17.5%という結果が得られた。要約スコアの適用により、要約率が70%と90%の場合に、baselineと比べて若干の誤り率削減を果たしたが、その効果は極わずかである。さらに、今回の実験では日本語としてもっともらしい要約文が必ずしも、音声認識に効果があるわけではなかった。しかしながら、要約文の言語モデルが粗製なものであるにも関わらず、認識率に若干の効果があったことから、要約文の言語モデルを改善することにより、さらに日本語としてもっともらしさを保持しつつ誤り率を削減できるものと考えられる。

5. むすび

話題語に着目したニュース音声の要約法について検討し、動的計画法に基づく要約アルゴリズムの提案、および、要約スコアによる音声認識への寄与を調べた。要約文として有望な結果が得られたが、原文の文意を損ねてしまうことがあり、さらに改善が必要である。さらに、要約文の日本語らしさと要約スコアによる認識性能への貢献の両立も、現段階では難しい。

今後は、これらの問題を解決するため、要約文の言語モデルの精度を向上させるべく、疑似要約文を含めて学習テキストについて検討を重ねていきたい。

参考文献

- [1] 岩崎淳 他, “ニュース音声からの話題抽出法の検討”, 音学秋季講論, 1-1-14, 1998.
- [2] 原田隆史 他, “抄録からの主題文の自動抽出”, 情報学基礎 29-3, 1993.
- [3] 仲尾由雄, “見出しを利用した新聞・レポートからのダイジェスト情報の抽出”, 自然言語処理, 117-17, 1997.
- [4] T. Wakao et al., “Text Summarisation for Production of Closed-Caption TV Programs in Japanese”, computer processing of oriental language, Vol.12, No.1, 1998.
- [5] 堀貴明 他, “大語彙連続音声認識のための音素グラフに基づく仮説制限法の検討”, 情報処理学会論文誌, Vol.40, No.4, pp.1365-1373, 1999.