話題語と言語モデルを用いた音声自動要約法の検討

堀 智織　　古井 貞熙

東京工業大学 情報理工学研究科 計算工学専攻
〒 152-8550 目黒区大岡山 2-12-1
chiori@cs.titech.ac.jp

あらまし　本稿では，音声認識結果から要約文を生成する音声自動要約手法を提案する．本手法は，認識された各発話文から相対的に重要な単語を，文字数を基準とした特定の割合で抽出し，それらを接合することにより要約文を生成する．要約文として抽出される単語列は，単語重要度 (重要度スコア) と言語尤度 (言語スコア) に基づいて決定される．要約文の尤もらしさを示す要約スコアを，抽出された各単語の重要度スコアと単語列の言語スコアの累積スコアとして定義し，この要約スコアを最大とする単語列を求めるために動的計画法を適用する．更に，要約スコアを単語数に基づいて正規化することにより，同じ発話文から様々な要約率で要約文を生成して最良の要約文を選択することを可能にした．NHK のニュース音声を大語彙連続音声認識システムを用いて音声認識し自動要約を行った．音声認識結果を 60–70% の文字数に要約した結果，発話文の重要単語のうち 86%が要約文に保持され，かつ，72%の要約文の文意が，発話文の文意に同意ないし包含されていることが確認された．

キーワード　音声要約，単語重要度スコア，言語尤度

Automatic Speech Summarization Based on Word Significance and Linguistic Likelihood

Chiori HORI and Sadaoki FURUI

Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, 152-8550 Japan

Abstract　This paper proposes a new method of automatically summarizing speech by extracting a limited number of relatively important words from its automatic transcription according to a target compression ratio for the number of characters. To determine a word set to be extracted, we define a summarization score consisting of a topic score (significance measure) of words and a linguistic score (likelihood) of the word concatenation. A set of words maximizing the score is efficiently selected using a dynamic programming (DP) technique. In order to evaluate the summarization scores of the summarized sentences obtained from the same original sentence using the various summarizing ratio, a normalization factor is applied to the summarization score. Japanese broadcast news speech transcribed using a large vocabulary continuous speech recognition system was summarized. As a result 86% of important words in the original speech were correctly included in the summarizing sentences and 72% of the summarizing sentences could maintain the meanings of the original speech under the 60–70% summarization condition.

key words　speech summarization, word significance score, linguistic likelihood

1　はじめに

近年の大語彙連続音声認識 (LVCSR) 技術の進展に伴い，LVCSR システムを用いて認識された音声を実用的な場面で利用することへの期待が高まっている．放送音声への字幕付与や議事録作成を考えた場合，英語などの表音文字ではキー入力した文字がそのまま表記となるので，LVCSR を利用しなくてもプロのタイピストが人間の音声を発声と同時に実時間で書き起こすことが可能である．一方，平仮名・片仮名・漢字を組み合わせた表意文字である日本語は，平仮名で入力された文字を片仮名や漢字に変換する必要があり，プロのタイピストであっても正確に実時間で音声を書き起こすことは非常に困難である．このため，日本語における大語彙連続音声認識 (LVCSR) システムを用いた音声のテキスト化が強く求められている．しかし，LVCSR システムの認識結果をそのまま字幕や議事録に用いると，冗長な情報が多く内容が簡潔に表現されていないという問題がある．例えば，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，会議や講演，講義では発話内容の重要な箇所が簡潔に抜粋されることにより，自動検索などに用いるインデクシングや抄録として用いることができる．また，人間の自然発話には，言い淀みや言い直し,「えー」などに代表される間投詞といった不要な情報が多く含まれているため，音声の書き起こしにこれらの不要な情報がそのまま含まれてしまう．これらの問題を解決するため，本研究では音声から重要箇所を適宜抜粋した要約テキストを出力する音声自動要約システムの構築を検討した．この音声要約技術は，音声付きの画像への字幕自動付与，議事録や講義・講演などの抄録の自動作成，または既存の音声データへのインデクシングなどに応用できる．

これまで自然言語処理の分野で検討されてきたテキスト要約の手法は，重要語などに基づき，文集合から重要文を抽出し要約文とする手法であった．また，字幕付与を目的として，放送前の原稿を用いて重要文を抽出し，さらに丁寧な表現を簡潔な表現にするなどの単語の置き換えを行うという要約手法がとられた [1]．一方，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，音声認識された各発話文の文頭から文末に向かい，原文の文字数に対し特定の要約率で単語を抽出し接合することで新たな単語列を生成する．抽出すべき単語は，各単語の重要度スコアと単語の連接によって決定される言語スコアによって決定される．この手法は，各単語の重要度と単語間の連接を考慮することで，発話文中の情報の核となる重要な単語とその他の単語を言語的に尤もらしく連接させることができる．

2　音声要約のアプローチ

発話文毎に，一定の割合の文字数で相対的に重要な単語を抽出し，接合することで要約文を生成する．各単語の重要度スコアと単語の連接によって決定される言語スコアに基づき，要約文に抽出されるべき単語が決定される．具体的には，重要度スコアと言語スコアの累積スコアによって，生成された単語列の要約文らしさを表す要約スコアを定義し，要約スコアが最大となる単語列を最適な要約文として，元の発話文の中から動的計画法を用いて求める．さらに，同一の発話文から異なる要約率で生成された要約文の間で，要約文としての尤もらしさを評価するため，要約スコアを単語数に基づき補正した．この手法は，重要な箇所を抜粋できるだけでなく，発話に含まれる冗長な情報の削除が可能となる特徴がある．

2.1　要約スコアの定義

要約スコアは，各単語の単語重要度スコアと単語列の連接に対する言語スコアによって次式のように定義される．N 個の単語からなる認識単語列 $W = w_1, w_2, \ldots, w_N$ から要約文として $M(M < N)$ 個の単語を抽出し接合した単語列 $V = v_1, v_2, \ldots, v_M$ の要約スコアは次式によって示される．

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

ただし，言語スコアには単語 trigram $P(v_m|v_{m-2}v_{m-1})$ を用いる．また，人間の被験者による先の実験 [3] で，話題語として選択された単語の多くが名詞であったことから，名詞の単語重要度スコア $I(v_m)$ には話題語らしさを表す話題語スコアを用いる．話題語スコアには，先の実験で最も精度の高かった以下の尺度を用いる．但し，

$$I(w_i) = g_i \log \frac{G_A}{G_i} \quad (2)$$

w_i :　ニュース音声中の名詞
g_i :　(注目している特定の) ニュース音声中の名詞 w_i の頻度
G_i :　ニュース原稿コーパス中の名詞 w_i の頻度
G_A :　ニュース原稿コーパス中の総名詞数 $G_i(= \sum_i G_i)$

一方，名詞以外の単語に対する単語重要度スコアには一定値を与えるものとする．

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，小さい時は単語の連接の日本語としての

尤もらしさに重点が置かれる．ここでは，実験的に決めた値を用いた。単語列より抜き出された部分単語列を $V = v_1, v_2, \ldots, v_M (M < N)$ とするとき，要約処理は (1) 式を最大にする $\hat{V}$ を求める問題となる．この問題は以下のように，動的計画法用いて解くことができる．

2.2 動的計画法による音声要約

N 単語からなる認識結果の単語列 $W = w_1, w_2, \ldots, w_N$ から，$M (M < N)$ 単語からなる単語列 $V = v_1, v_2, \ldots, v_M$ を抽出し，要約文候補の中から式 (1) で与えられる要約スコアを最大とする要約文を決定するアルゴリズムを以下に示す．

(1) 記号と変数の定義

<s> ：文頭記号
</s> ：文末記号
$g(m,l,n)$ ：局所最適スコア

(m 単語で構成され，<s>で始まり w_l,w_n で終る部分単語列 <s>$, \ldots, w_l, w_n$ の要約スコア，但し，$(0 \leq l < n \leq N)$)

$B(m,l,n)$ ：バックポインタ

(2) 初期設定

$$g(1,0,n) = \begin{cases} \log P(w_n|\texttt{<s>}) + \lambda I(w_n) \\ \qquad if\ 1 \leq n \leq (N-M+1) \\ -\infty \qquad otherwise \end{cases}$$

(3) 漸化式計算

for $m = 2$ to M
  for $n = m$ to $N - m + 1$
    for $l = m - 1$ to $n - 1$

$$g(m,l,n) = \max_{k<l}\{g(m-1,k,l) + \log P(w_n|w_k w_l) + \lambda I(w_n)\}$$

$$B(m,l,n) = \operatorname*{argmax}_{k<l}\{g(m-1,k,l) + \log P(w_n|w_k w_l) + \lambda I(w_n)\}$$

(4) 最適パスの選択

$$S(\hat{V}) = \max_{\substack{N-M<n\leq N \\ N-M-1<l\leq N-1}} g(M,l,n) + \log P(\texttt{</s>}|w_l w_n)$$

$$(\hat{n},\hat{l}) = \operatorname*{argmax}_{\substack{N-M<n\leq N \\ N-M-1<l\leq N-1}} g(M,l,n) + \log P(\texttt{</s>}|w_l w_n)$$

(5) トレースバック

for $m = M$ to 1

$v_m = w_{\hat{n}}$
$l' = B(m,\hat{l},\hat{n})$
$\hat{n} = \hat{l}$
$\hat{l} = l'$

動的計画法の処理過程を図 1 に示す．

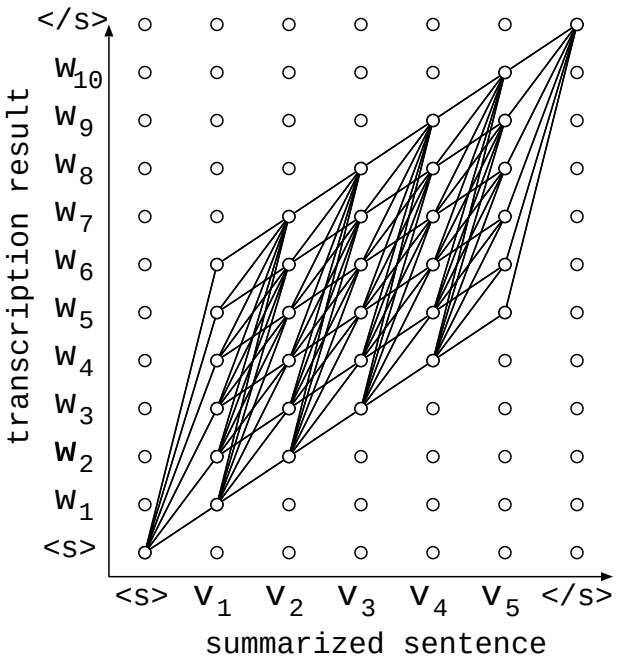

図 1: 音声要約のための動的計画法の計算領域．

式 (1) で求めれる要約スコアは，要約文として抽出される単語列の単語数と共に単調に変化する．そのため，同一の認識文から抽出した要約率の異なる要約文の要約スコアを比較することができない．そこで，要約率による要約スコアの変化を単語数により正規化し，この正規化要約スコアが最大となる要約率の文を最良の要約文として選択することにする．正規化要約スコア $S^*(M)$ は次のように計算する．

$$S^*(M) = \bar{S}(M) - \bar{S}_{adj}(M) \tag{3}$$

ここで，

$$\bar{S}(M) = S(M)/M$$

$$\bar{S}_{adj}(M) = \frac{\bar{S}(M_{max}) - \bar{S}(M_{min})}{M_{max} - M_{min}}(M - M_{min})$$

$S(M)$ ： 動的計画法によって得られる単語数 M の要約文の要約スコア
M ： 要約文の単語数 $(M_{min} \leq M \leq M_{max})$
M_{min} ： 要約文の単語数の下限
M_{max} ： 要約文の単語数の上限

3　音声認識システム

3.1　特徴抽出

音声データを 12kHz，16bit でディジタル化し，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．

3.2　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，４混合分布) 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 (約 20 時間) を用いた．

3.3　言語モデル

言語モデルとして単語 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．

言語モデルを作成する際，文脈に依存した読みの制約を考慮するために，同じ漢字表記であっても読みが異なる場合は別の単語として扱っている．さらに，先の検討 [4] において不適切な品詞のつながりが要約文の文意を損ねるケースがあったことから，品詞情報も単語に組み込み，同じ表記，読みであっても品詞が異なる場合は別の単語として扱うようにした．

3.4　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 [6] 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．

4　音声要約実験

4.1　評価データ

提案手法を評価するために 1996 年 6 月の放送ニュース音声を用いた．評価セットは５名のアナウンサーが発声した計 48 発話で, 人手によって文単位に分割されている．未知語率は 20000 語に対して 1.8%で，音声認識結果は単語正解精度 85.4%である．

要約文の評価は，このうち単語正解精度が 90%以上の文 21 発話を選んで行った．要約文は，要約率を認識結果の文字数に基づいて 60–70%の間で変化させ，2.2 節で述べた正規化要約スコアが最大となる要約文を要約結果として採用した．70%は，発話された文の文字数に対し人間が字幕として容易に読むことのできる適正な文字数の割合であると考えられている．

4.2　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 trigram 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意する．

本来，要約文の trigram 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，先の検討 [4] では意図的に冗長な表現を排除したニューステキストを作成し，これを元のニュース原稿と混合して用いていた．しかしながら，機械的に冗長な表現を排除すると，日本語として正しくない系列が生じることがあり，不自然な要約文が生成されることがあった．そこで，より要約文の言語モデルの作成に相応しいテキストとして，新聞記事を採用することにした．新聞記事はニュース原稿に比べてよりコンパクトで簡潔な表現を含むことから，要約文の言語モデルを構築するのに適したテキストと考えられる．新聞記事コーパスとして，評価用音声と同時期の毎日新聞 1996 年の記事 (約 27M 単語) を用いた．比較のために，認識時に使用したものと同じニュース原稿に基づく trigram を用いて要約文を生成する実験も行った．

4.3　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章で述べた手法を用いて生成した．さらに, 評価音声に対し人間が書き起こしたテキストに対しても, 同様の手法で要約文を生成し, 比較を行った．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が抽出されているか，および，原文に対して要約文の文意が保たれているか，という２点に関して評価した．

(1) 抽出されるべき単語

８人の被験者が，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．書き起こしテキストから生成された要約文と比較すると，わずかではあるが，認識誤りによる抽出洩れが生じた．

(2) 要約文の文意

要約文の文意について８人の被験者による以下の３段階評価を行った．

(1) 原文と同意である．
(2) 原文の文意に包含されている．
(3) 原文の文意とは異なる．

評価結果を表 1 にまとめる．音声認識結果を要約対象とした場合，要約用言語モデルをニュース原稿，新聞記事のいずれから作成しても，全体の 72%の要約文が原文の文意に同意ないし包含されていることが分かる．音声認識結果に対しては，要約文作成に用いた言語モデルの違いによる差は認められなかったが，書き起こしテキストに対する要約では，新聞記事を用いた方が原文と文意が「異なる」ものが 10%削減されることが分かる．

表 1: 要約文の文意に対する評価

言語モデル	要約対象	同意	包含	異なる
ニュース原稿	書き起こし	33%	43%	24%
	音声認識結果	25%	47%	28%
毎日新聞	書き起こし	39%	47%	14%
	音声認識結果	23%	49%	28%

4.4 要約文例

新聞記事を用いて作成した要約用言語モデルを用いて音声認識結果を要約した要約文の中から,「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の例を原文と併せて以下に示す．比較として，書き起こし正解文に対する要約も付す．

(1) 原文と同意
書き起こし文
【原文】
日教組は去年の定期大会で学習指導要領を容認するなど文部省との協調路線を柱とする運動方針を決め大幅な路線転換を行いました
【要約文】
日教組定期大会で学習指導要領を容認するなど，文部省との協調路線を柱とする方針を決め，路線転換
音声認識結果
【原文】
日教組は去年の定期大会で学習指導要領を容認するなど文部省との協調路線を柱とする運動方針を決め大幅な路線転換を行いました
【要約文】
日教組定期大会で学習指導要領を容認する文部省との協調路線を柱とする方針を決め，路線転換

(2) 原文に包含
書き起こし正解文
【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】
与党は住専処理法案など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，法案の攻防が続く見通し
音声認識結果
【原文】
これを受けて与党 側が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あす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】
これを受けて与党が住専処理法案などの採決を目指すことにして新進党は，採決には抵抗する構えで，法案の攻防が続く見通し

(3) 原文と異なる
書き起こし正解文
【原文】
その上で菅厚生大臣は厚生省が一丸となって諮問に向けて努力し早い時期に審議会の答申を頂き今月十九日の国会の会期末までになんとか法案として提出したいと述べ厚生省としてはあくまでも今の国会の会期中の法案提出を目指す立場から今月六日に制度案の諮問を行いたいという考えを示しました
【要約文】
菅厚生大臣は，厚生省が諮問審議会の答申を今月十九日の国会会期末までに法案提出したいと述べ，厚生省としては，今の国会の会期中の法案提出を目

指す今月六日に，制度の諮問を行いたいという考えを示し

音声認識結果

【原文】

その上で菅厚生大臣は厚生省が一丸となって諮問に向けて努力し早い時期に審議会の答申も板垣今月十九日の国会の会期末までになんとか法案として提出したいと述べ厚生省としてはあくまでも今の国会の会期中の法案提出を目指す立場から今月六日に生徒の指導を行いたいという考えを示しました

【要約文】

厚生省が一丸となって努力し，今月十九日の国会の会期末までに法案提出したいと述べ，厚生省としては今の国会の会期中の法案提出を目指す立場から，今月六日に指導を行いたいという考えを示し

5 まとめ

話題語と言語尤度に基づく要約スコアを最大化する規準で，動的計画法を用いる自動音声要約手法を提案した．日本語として不適切な単語の連接が生じるのを避けるために，より良い要約文の言語モデルについても検討を行った．本手法は，原文の文意を含みつつ効果的に単語数を削減する有効な手法であることが確認された．ニュース音声を認識し，文字数について 60–70%の要約率で要約文を生成した結果，原文に出現した重要語のうち 86%が要約文中に含まれ，全体の 72%が原文の文意を内包していることが確認された．要約用言語モデルを作成するコーパスとして，ニュース原稿よりも新聞記事を用いた場合に，書き起こしテキストにおいて顕著な改善が見られた．また，誤りを含む音声認識結果を用いる場合でも，単語重要度スコアにより一部の認識誤りが削除され，正解文の文意を保持した要約文を生成することができた．今後は，誤って要約される文章の割合を削減するために，単語の意味情報など，より上位の知識を利用することが必要と考えられる．さらに，具体的な要約文の例を集めて，その解析を進めていきたい。

参考文献

[1] T. Wakao et al., Computer Processing of Oriental Language, Vol.12, No.1,1998.

[2] S. Furui et al., Proc. DARPA Broadcast News Transcription and Understanding Workshop,pp.144-149, 1998.

[3] 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.

[4] 堀 智織 他, “話題語に着目したニュース音声の要約法の検討”, 音学秋季講論,?-?-?, 1999.

[5] C. Hori and S. Furui, Proc. 1999 Japan-China Symposium on Advanced Information Technology, pp. 75-82, 1999.

[6] 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.