

1. はじめに

これまで、音声自動要約手法として、音声認識された各発話文から相対的に重要な単語を、特定の要約率(認識結果の文字数に対する要約文の文字数の割合)で抜き出し、それらを接合することによって要約文を生成する音声自動要約の枠組を提案してきた。この発話単位の自動要約手法は、要約文の尤もらしさを示す要約スコアを最大とする部分単語列を最適な自動要約文として動的計画法により決定する。要約スコアは、自動要約文に抽出された各単語の単語重要度(重要度スコア)、単語連鎖の言語尤度(言語スコア)[1][2]、音声認識時における各単語の音響的、言語的信頼度(信頼度スコア)[3]、および原文の係り受け構造に基づく単語間遷移確率(単語間遷移スコア)[4]の累積スコアに基づき定義される。さらに、この発話単位の要約手法を複数発話から構成される主題のある音声の要約に拡張し、全体として特定の要約率となるよう、各発話文を可変的な要約率で要約する複数発話自動要約手法を提案した[4]。

本稿では、この複数発話自動要約手法を用いてニュース音声の記事単位で自動要約した結果を報告する。さらに、自動要約文の評価方法として正解要約文単語ネットワークを用いた新しい評価方法を提案する。

2. 音声自動要約手法

2.1 要約スコアの定義

要約スコアは、単語重要度スコア I と言語スコア L 、信頼度スコア C 、および、単語遷移スコアに基づき、次式のように定義する。 N 個の単語からなる認識単語列 $W = w_1, w_2, \dots, w_N$ 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 $V = v_1, v_2, \dots, v_M$ の要約スコアは次式によって示される。

$$S(V) = \sum_{m=1}^M \{L(v_m | \dots v_{m-1}) + \lambda_I I(v_m) + \lambda_C C(v_m) + \lambda_T Tr(v_{m-1}, v_m)\} \quad (1)$$

但し、 $\lambda_I, \lambda_C, \lambda_T$ は各スコアのバランスをとるための重み係数である。

認識された単語列より抽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(1)式で表される要約スコアを最大にする \hat{V} を求める問題となる。

【単語重要度スコア】

単語重要度スコア $I(v_m)$ は、原文における相対的な単語の重要度を示すスコアである。本研究では、単語重要度スコアとして単語の出現頻度に基づく情報量を適用する。ただし、話題語に重み付けするため、名詞以外の単語には一定値を与える。さらに、冗長性を削減するため、2回目以降の名詞にも一定値を与える。

【言語スコア】

言語スコア $L(v_m | \dots v_{m-1})$ は、要約文内の単語連鎖の適正度を示すスコアである。本研究では、単語 trigram を用いる。

【信頼度スコア】

信頼度スコア $C(v_m)$ は、認識結果に含まれる認識誤りを要約文に抽出しないよう、音響的、言語的に信頼度の低い

単語に対しペナルティを与えるものである。デコーダから出力された単語グラフに付与された音響尤度および言語尤度に基づく各単語に対する事後確率の対数値を、信頼度スコアとして用いる。

【単語間遷移スコア】

単語間遷移スコア $Tr(v_{m-1}, v_m)$ は、要約文内の単語連鎖が原文において係り受け関係にあるか否かを示す単語間遷移確率の対数値で定義され、係り受け関係にない単語連鎖にペナルティを与えるものである。文節単位の係り受け SCFG(SDCFG:Stochastic Dependency Context Free Grammar) [5] に基づき、文節境界を越える遷移を考慮した単語間遷移確率を単語間遷移スコアと定義する。文節内の単語間での遷移確率は、前方から後方の単語に係るという正規文法を定義し、係り受け関係のある単語間の遷移確率を 1、係り受け関係のない単語間の遷移確率を 0 とする。また、文節境界を越える単語間の遷移確率は、各単語が属する文節の係り受け構造に基づき単語間遷移を規定する。

2.2 動的計画法による複数発話音声要約

発話単位の要約手法を用い、複数の文を一文として扱うことにより複数発話の要約手法に拡張する。但し、文境界の遷移では単語間遷移スコアを用いず、さらに前文の文末記号と後続する文の文頭記号を必ず遷移すると規定する。要約スコアには単語重要度スコアが含まれているため、主題のある複数の発話を対象とするこの複数発話要約手法では、重要な情報を多く含む文は文長が長く、そうでない文は短くなるか完全に削除される。この手法は従来の重要文抽出に基づく要約手法と、これまで提案してきた発話単位の要約手法を統合した要約手法であり、限られた文字数でより情報量の多い要約文を生成することが可能となる。

J 個の発話文 S_1, \dots, S_J ($S_j = w_{j1}, w_{j2}, \dots, w_{jN_j}$) より、 M ($M < \sum_j N_j$) 単語からなる部分単語列 $V = v_1, v_2, \dots, v_M$ を、式(1)で与えられる要約スコアが最大となるように決定する。動的計画法の処理過程を図1に示す。

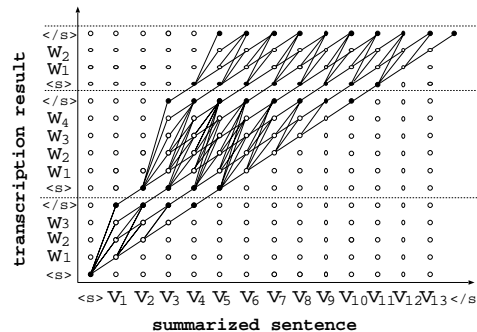


図 1. 複数発話の音声要約のための動的計画法の計算領域

3. 正解要約文単語ネットワークに基づく自動要約法の評価

人間によって作成された正解要約文に関して、日本語としての適正度、原文の文意の保持という点から、要約の良さを定量的に評価する手法として、これまで単語連鎖適合率[3]を用いてきた。この単語連鎖適合率は、被験者間で

* An approach to automatic speech summarization for multiple utterances with consistent meanings.

正解要約文が異なり、さらに被験者による正解要約文が全ての正解要約文を網羅していないということから、長さの異なる部分単語連鎖による適合率を用いることにより、全ての可能性のある正解要約文に対する評価を行うとした尺度である。一方、本稿で提案する正解要約文単語ネットワークは、被験者の作成した正解要約文を単語間の連鎖をまとめることにより、全ての可能性のある正解要約文の単語連鎖を近似的に網羅している。自動要約文は、このネットワーク上で最も近い要約文に対して単語正解精度により一元的に評価できる。図2に正解要約文単語ネットワークの例を示す。

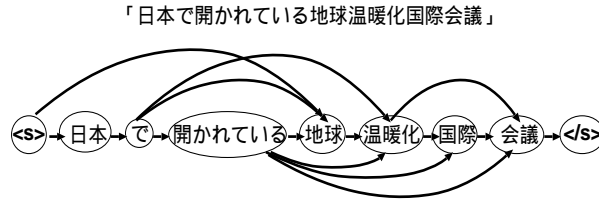


図2. 正解要約文単語ネットワークの例

4. 評価実験

4.1 実験条件

女性話者1名による放送ニュース音声で大語彙連続音声認識システムにより音声認識し、平均単語誤り率90%以上の平均5発話で構成された5つのニュースの書き起こし文 (TRS) および音声認識結果 (REC) を用いて、提案手法により20-30%の要約率で自動要約を行った。生成された自動要約文を、12人の被験者の作成した正解要約文に基づき単語連鎖適合率および単語ネットワークに基づき評価を行った。比較のため、SDCFGに基づく単語遷移スコアを適用しない要約文 (NOSDCFG) も生成した。各被験者の要約文をその他の11人の要約文を正解として評価したSUB_TRSを目標値とした。

4.2 音声認識システムの構成

【特徴抽出】

音声データを16kHz, 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12次元のメルケプストラムおよびメルケプストラム(計25次元)を抽出する。さらに発話毎にケプストラム平均正規化を行う。

【音響モデル】

不特定話者音素文脈依存HMM(1012状態), 混合ガウス分布のIPAの女性モデルを用い、評価話者の985発話(約2時間)を用いて最尤推定話者適応を行った。

【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JUMANを用いて形態素に分解し、「単語+読み+品詞」を一つの形態素としてbigramとtrigramを学習した。

【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。

4.3 要約処理部の構成

【単語重要度スコア】

音声認識用言語モデルを学習した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50万文を用い、全文における各単語の出現頻度に基づき重要度スコアを求める。

【要約言語モデル】

要約対象であるニュース音声と同時期で、要約文に求められる簡潔な表現が多く含まれている新聞テキストを用いて要約用言語モデルの学習を行った。3年分の毎日新聞(1996-1998)のテキストを用い、音声認識に用いた言語モデルの20000語彙を用いて単語trigramを学習した。

【文節単位のSDCFG】

毎日新聞約4万文の構文解析済みの京大テキストコーパスを用い、構文木制約付きのInside-Outsideアルゴリズムを用いて、SDCFGのパラメータの推定を行った。但し、非終端記号数は100とした。

4.4 自動要約文の評価結果

音声自動要約の結果を示す。単語連鎖適合率による評価結果を図3に示し、正解要約文単語ネットワークによる評価結果を表1に示す。

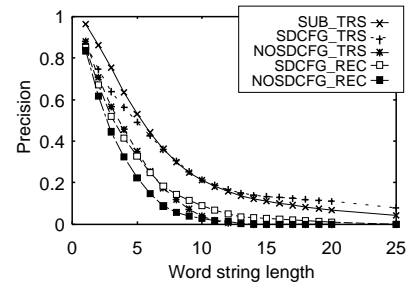


図3. 主題のある複数発話要約文の単語連鎖適合率と単語連鎖長の関係

表1. 正解要約文単語ネットワークに基づく単語正解精度

	TRS	REC
SUB	82.3%	-
SDCFG	74.6%	70.7%
NOSDCFG	69.3%	65.0%

自動要約文 (SDCFG) は、単語連鎖適合率では被験者の作成した要約文 (SUB_TRS) に匹敵しているが、正解要約文単語ネットワークに基づく単語正解精度ではSUB_TRSに到達していない。これは、被験者による正解要約文間の相違が大きいことによるものである。複数文を自動要約する場合、発話単位の要約に比べ単語の選択肢が多くなり、十分な被験者による正解要約文を得ることは困難である。今後、複数文の自動要約結果を用いたタスクによる評価法の検討を更に行う必要がある。

5. まとめ

本稿では、複数発話自動要約手法を用いてニュース音声を記事単位で自動要約し、正解要約文単語ネットワークを用いて評価を行った。SDCFGを用いることにより複数発話においても文脈を考慮した自動要約文を生成できた。

参考文献

- [1] 堀, 古井, 情処学研報, 99-SLP-29, pp.103-108(1999). 信学技報, SP99-110, pp.103-108(1999).
- [2] Hori and Furui, Proc. ICASSP2000, Istanbul, Vol.3, pp.1579-1582(2000).
- [3] Hori and Furui, Proc. ICSLP2000, Baijing, Vol.4, pp.326-329(2000).
- [4] 堀, 古井, 信学技報, SP2000-95-116, pp.127-132(2000).
- [5] 堀, 加藤, 伊藤, 好田, 信学論, Vol.J83-D-II, No.11, pp.2407-2417(2000).