

1 はじめに

これまで我々は、単語抽出に基づく音声自動要約システム [1] を提案してきた。本稿では、この音声自動要約システムを用い、講演録作成を目的として講演音声を自動要約し評価した結果を報告する。複数の発話で構成された音声を要約する手法として、各発話を可能性のある全ての要約率で要約後、全体として目的の要約率となるよう各文の要約文を組み合わせる 2 段 DP による要約手法を提案する。

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 T_r 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、 λ_I 、 λ_C 、 λ_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。

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。

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。発話の重要な情報を担う話題語は名詞であることが多いため、名詞の単語重要度スコアには情報量を適用し、名詞以外の単語のスコアは一定値とした。ただし、名詞には下記に示す動作を表す名詞が含まれている。

「検討する」 「検討(名詞)+する(動詞)」

さらに、冗長性を削減するため、2 回目以降の名詞のスコアも一定値とした。

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 $Tr(v_{m-1}, v_m)$ 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) [2] 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 [3]。

2.2 動的計画法による複数発話音声要約

これまで、複数発話要約手法として、文間でのスコアの適用を制約することにより、全文をまとめて自動要約する手法を提案してきた [3]。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。これにより、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。この複数発話要約手法は、従来の重要文抽出に基づく要約手法と、発話単位の要約手法を統合した手法である。

しかし、この手法は、要約対象となる発話数の増加に伴い、単語の組み合わせの数が増加し、計算量が増大してしまう。本研究では、複数発話の要約手法として、2 段 DP による要約手法を提案する。本手法は、第一段階として、各発話要約手法により可能な全ての要約率で各文を要約する。さらに、第二段階として、全体が目的の要約率となるよう各文の要約文を組み合わせ、その中から要約スコアが最大となる組み合わせを動的計画法により決定する。この手法は、単語単位で複数文を要約する処理に比べ、計算量の増加を抑制しつつ、全体に渡る重要度により各文を伸縮するという同様の効果を実現できる。

N 個の発話文 S_1, \dots, S_N ($S_N = w_{j_1}, w_{j_2}, \dots, w_{j_{N_j}}$) 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、要約スコアが最大となるように決定する。動的計画法の第 2 段の処理過程の例を図 1 に示す。

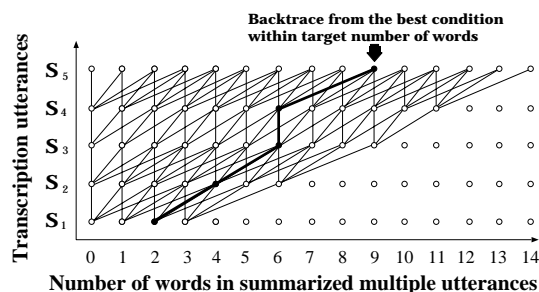


図 1. 複数発話の音声要約のための動的計画法の計算領域の例

* Automatic speech summarization of presentation utterances for making abstracts.

3 評価実験

被験者の作成した正解要約文に基づき作成された単語ネットワーク上で、自動要約文自身に最も近い単語列を正解として、要約正解精度により自動要約文を評価する [3] .

3.1 要約処理部の構成

【要約言語モデル】

話し言葉コーパスの書き起こしを、自動要約における前処理を施して論説調の表現に変換し、さらにテストセット以外の講演の予稿集を加えて単語 trigram を学習した .

【単語重要度スコア】

音声認識用言語モデルを学習した講演の書き起こし、テストセット以外の講演の予稿集、WWW 上の講演録、NHK のニュース原稿、毎日新聞および「音声情報処理」のテキストを用い、出現した全約 120k 種類の単語の各々の出現頻度に基づき重要度スコアを計算した .

【文節単位の係り受け SCFG】

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、係り受け SCFG 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 .

3.2 要約実験条件

話し言葉コーパス中の男性話者 (AS99SEP097)1 名による講演音声の書き起こし (TRS) および大語彙連続音声認識システム [4] による単語正解精度約 70% の音声認識結果 (REC) を、提案手法を用いて (70-)80% と (40-)50% の要約率で自動要約を行った .

書き起こし (TRS) については単語重要度スコア (I)、言語スコア (L) を単独に用いた場合、単語重要度スコアと言語スコアを組み合わせた場合 (IL)、さらにそれらに単語間遷移スコアを組み合わせた場合 ($ILLT$)、全 4 種類の自動要約文を生成した . さらに、認識結果 (REC) については、信頼度スコア (C) を組み合わせ、全 6 種類の自動要約文を生成した . ただし、単語重要度スコア、信頼度スコア、単語間遷移スコアの重み係数 λ_I 、 λ_C 、および λ_T は、実験的に求めた最適値を用いた . 生成された自動要約文を、9 人の被験者の作成した正解要約文に基づき、要約正解精度によって評価した .

さらに、提案手法の有効性を検証するため、自動要約文と等しい要約率で単語をランダムに抽出した要約文 (RDM) に対して評価を行った . さらに、被験者各 9 人の被験者の正解要約文を他の 8 人の正解要約文で作成した正解要約文単語グラフに基づき評価した平均単語正解精度 (SUB) を示す .

3.3 評価結果

実験結果を図 2, 3 に示す . 全ての自動要約条件において、ランダムに単語を抽出した場合と比較して、要約正解精度が有意に高くなることが示された . 書き起こし (TRS) を 80% に要約した場合を除いて、単語重要度単独の I に比べ、言語尤度単独の L の精度が高く、さらに、両者を組み合わせた IL の精度が高いことが示された . L が I よりも低い場合、 L を組み合わせることによる効果は無い . 単語間遷移スコア、および信頼度スコアを組み合わせた効果は、要約率 50% における認識結果 (REC) でわずかに示さ

れたに留まっている . 単語間遷移スコアは、新聞記事の係り受け関係を学習した SDCFG に基づいていることから、ニュース音声の要約においては有効であったが、話題も構造も異なる講演音声では有効性は示されなかった . また、書き起こし (TRS) と認識結果 (REC) の要約精度の差が大きいことから、認識誤りにより要約性能が劣化していると考えられる . 80% 要約率では被験者の作成した正解要約文の精度に近い値となっているが、50% 要約率では正解精度には至っていない . さらに、一人の被験者により音声認識結果を要約した結果は、自動要約と同程度であった .

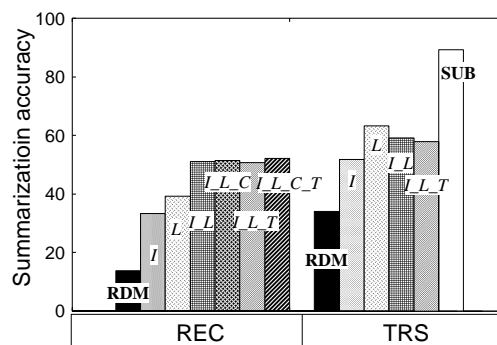


図 2. 要約率 50% 時における要約正解精度

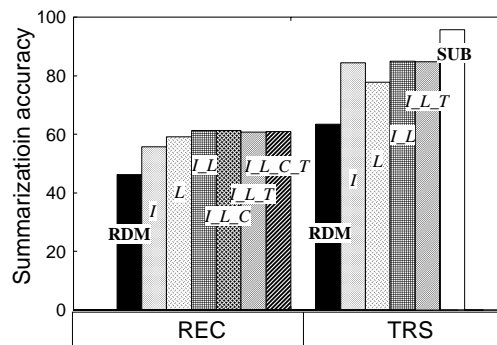


図 3. 要約率 80% 時における要約正解精度

4 まとめ

本稿では、講演音声の講演録作成を目的として 2 段 DP により複数の発話で構成された講演音声の自動要約を行った . 今後、自然発話音声の自動要約文を作成するため、自然発話を対象とした音声認識性能の向上が必須であり、さらに、係り受け関係の推定精度を向上させるためには、話し言葉コーパスを構文解析したデータベースの整備が不可欠である .

参考文献

- [1] 堀, 古井, 信学技報, SP2000-95-116, pp.127-132(2000).
- [2] 堀, 加藤, 伊藤, 好田, 信学論, Vol.J83-D-II, No.11, pp.2407-2417(2000).
- [3] 堀, 古井, 音講論, Vol.1, 2-3-20, pp.93-94(2001-3).
- [4] 篠崎, 古井, 音講論, Vol.1, 1-3-14, pp.31-32(2000-1).