音声自動要約手法の英語ニュース音声への適用 *

◎堀 智織　古井 貞熙　(東工大)

1　はじめに

これまで、日本語を対象として，音声認識結果から，文字数または単語数を基準とした特定の割合 (要約率) で，要約文として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の抽出を行う自動要約手法を提案してきた [1][2]．本稿では，英語のニュース音声を対象として，音声認識結果および書き起こし文の自動要約を試みた結果を報告する．英語の係り受け構造は日本語と異なることから，原文の係り受け関係に基づく単語間遷移確率を推定するモデルを拡張した [3].

2　音声自動要約手法

これまで我々は，話題語を中心として，音声認識による認識誤りを排除しつつ，原文の文意をできる限り保持し，言語的に尤もらしくなる部分単語列を抽出する音声自動要約手法を提案した．本手法では，特定の要約率に基づいて，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，動的計画法により決定する．

単語抽出により生成された要約文の適正を示すスコアとして要約スコアを定義する．要約スコアは，単語重要度スコア I，言語スコア L，信頼度スコア C，および，単語間遷移スコア T の重みつきの和で表す．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．

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

ここで，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 $V = v_1, v_2, \ldots, v_M (M < N)$ 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．

単語間遷移スコア T は，原文の単語間の係り受け関係に基づく単語間遷移確率によって定義する．単語間の係り受け関係は，SDCFG(Stochastic Dependency Context Free Grammar) の確率を用いて推定する．英語には，前方から後方への係り受けと後方から前方への係り受けが存在する．英語の SDCFG は，$\alpha \to \beta\alpha$，$\alpha \to \alpha\beta$，および，$\alpha \to w$ の規則からなる．ここで，α,β は任意の非終端記号，w は終端記号 (単語) を表す．これらの規則が適用される確率を基に，Inside-Outside 確率を計算し，二つの単語が係り受け関係にある確率 (係り受け確率) を求める．

図 1 において，w_m と w_l の係り受け確率を $d(w_m, w_l, i, k, j)$ と定義する．ここで，係る単語 w_m を「修飾語」，係られる単語 w_l を「主辞」と呼ぶ．要約文では，直接係り受け関係がある修飾語と主辞の間，および，修飾語と主辞を修飾している単語間で連接が可能であることから，w_m と w_n の単語間遷移確率を，w_m と w_n の係り受け確率と，w_m と $w_{n+1} \ldots w_l$ の各単語間の係り受け確率の総和で定義する．これにより，w_m と w_n の単語間遷移スコアは次式のように単語間遷移確率の対数値で定義される．

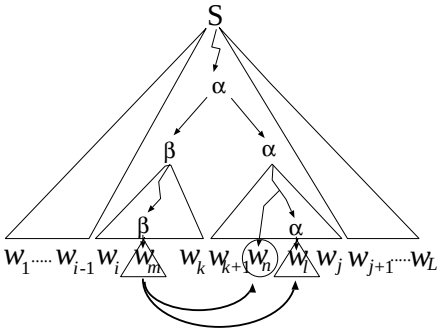

図 1. SDCFG に基づく係り受け構造

$$T(w_m, w_n) = \log \sum_{i=1}^{m} \sum_{k=m}^{n-1} \sum_{j=n}^{L} \sum_{l=n}^{j} d(w_m, w_l, i, k, j) \quad (2)$$

複数発話を要約する手法として，2 段 DP による要約手法を適用した [4]．この手法は，各発話文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各文の要約文の組み合わせから，要約スコアを最大とする組み合わせを決定する．

3　評価実験

3.1　実験条件

NIST による Topic Detection and Tracking(TDT) のタスクから，CNN の 5 ニュースを選び，音声認識結果と音声認識誤りを含まない正解書き起こし文を自動要約の対象とした．単語に品詞情報を付加して各モデルを学習し，評価時にも品詞情報を付加した．品詞付けには Brill tagger を用いた [5]．5 ニュース中の 50 発話文について，70%と 40%の要約率で各文要約を行った．さらに，5 ニュースをニュース単位で，70%と 40%の要約率で複数文要約を行った．但し，要約率は原文の単語数に対する自動要約文の単語数の割合によって定義される．生成された自動要約文を，英語を母国語とする 17 人の被験者の作成した正解要約文の単語ネットワーク [3] に基づき評価を行った．

3.2　音声認識部の構成

英語のニュース音声を認識するため，JRTk(Janus Speech Recognition Toolkit) [6] を用いて．特徴抽出は，13 次元のメルケプストラム (MFCC) を抽出し，7 フレームを 1 セグメントとして特徴ベクトルの単位とし，各セグメントを LDA(Linear Discriminant Analysis) に基づいて 42 次元に圧縮する．音響モデルは，2000 個のガウス分布コードブックを 6000 状態の各混合重み分布で共有した，計

∗ Application of automatic speech summarization technique to English broadcast news speech.
By Chiori Hori and Sadaoki Furui(Tokyo Institute of Technology)

105k のガウス分布を持つ quinphone HMM を用いる．学習データには Wall Street Journal(WSJ)，English Spontaneous Scheduling Task(ESST)，Broadcast News(BN)，Crossfire and Newshour TV news show を用いている．言語モデルとして，BN コーパスから学習した語彙サイズ 40k の単語 trigram を用いる．デコーダは，第 1 パスでは triphone と bigram を用いて木構造辞書に基づく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．第 2 パスでは単語グラフ上の単語からフラットな辞書を構成し，quinphone と trigram を用いてビームサーチを行い単語グラフを再構成する．第 3 パスでは，単語グラフを最小化した後で trigram を用いてリスコアを行い最終的な認識結果を得る．

3.3 要約処理部の構成

単語重要度，要約用言語モデル，および単語間遷移確率に適用する SDCFG は，Penn Treebank コーパスに含まれる Wall Street Journal と BROWN コーパス 10681 文 (約 3.5M 単語) を用いて算出した [7]．SDCFG は，Penn Treebank に付与された非終端記号を適用せず，括弧付きのコーパスとして非終端記号数のみを固定して学習した．ただし，非終端記号数は 100 である．

3.4 評価結果

図 2 に各文要約，図 3 に複数文要約の正解要約文単語ネットワークに基づく要約正解精度を示す．音声認識の単語正解精度は，各文要約に用いた文が約 81.4%，複数文要約に用いた文が 78.4%である．

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，上限として，被験者各 17 人の被験者の正解要約文を，他の 16 人の正解要約文で作成した要約文単語ネットワークに基づき評価した要約正解精度 (SUB) を示す．

各文要約，複数文要約のどちらにおいても，単語重要度 I に比べ言語尤度を組み合わせた I_L の精度が高い．さらに，単語間遷移スコアを組み合わせた I_L_T の要約精度が最も高いことが示されている．音声認識結果に対する自動要約では，信頼度スコア C を組み合わせることにより，さらに改善されている．以上の点から，原文の係り受け関係を考慮し，認識誤りの自動要約文への抽出を抑制することにより，誤要約が削減されたものと考えられる．ただし，被験者の作成した正解要約文の精度にはまだ至っていない．

提案手法による自動要約文に抽出された誤認識単語の数と，その誤認識単語が含まれる自動要約文の数を表 1 に示す．表 1 より，提案手法が，誤認識された単語が含まれていることによる誤要約を削減できていることが分かる．言語スコアにより，音声認識誤りが削減されているのは，文脈に整合しない認識誤りが排除されたことに因るものと考えられる．さらに，信頼度スコアにより，顕著に認識誤りが排除されていることが分かる．

表 1. 要約文中の認識誤りと認識誤りを含む要約文数

	各文要約		複数文要約	
要約前	180 単語 (45 文)		326 単語 (94 文)	
要約率	40%	70%	40%	70%
I	42 (27)	111 (40)	99 (56)	199 (71)
I_L	44 (28)	87 (37)	86 (53)	166 (69)
I_L_C	23 (15)	49 (22)	34 (28)	82 (47)
I_L_T	46 (27)	84 (37)	90 (56)	173 (69)
$I\_L\_C\_T$	22 (13)	51 (24)	25 (17)	80 (47)
RDM	82 (30)	87 (21)	89 (45)	169 (65)

() 内は認識誤りが含まれる要約文の数を表す．

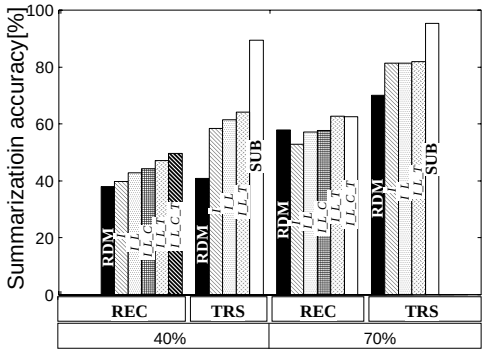

図 2. 各文要約の要約正解精度

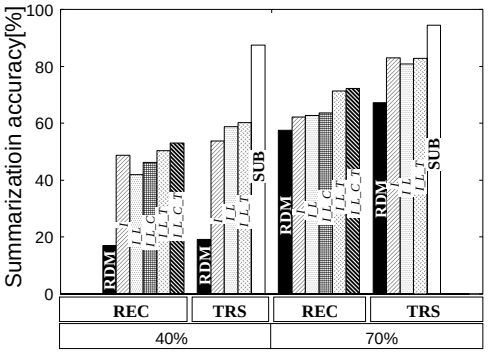

図 3. 複数文要約の要約正解精度

4 まとめ

英語のニュース音声の自動要約の結果，日本語のニュース音声の自動要約の結果と同様，提案手法が相対的に冗長な情報を削除し，重要な情報を抽出していることが示された．日本語のニュース音声の要約と比較して，英語のニュース音声の要約では信頼度スコアによる改善が大きい．これは，日本語のニュース音声の自動要約では，信頼度スコアは自動要約文中の認識誤りを削減する作用に留まっているのに対し，英語のニュース音声の自動要約では，さらに明確に発音された重要な単語を抽出する作用があるものと考えられる．

謝 辞

音声認識装置を提供し，英語の音声認識に貢献してくださった Carnegie Mellon University の Alex Waibel 教授，Rob Malkin 氏，Hua Yu 氏に感謝致します．正解要約文作成に協力してくださった Sheffield 大学の Yoshi Gotoh 氏に感謝致します．

参 考 文 献

[1] 堀，岩崎，古井, 音学秋季講論,Vol.1, 3-1-11, pp.117-118 (1999).

[2] C. Hori and S. Furui, Proc.ICASSP2000, Vol.3, pp.1579-1582 (2000).

[3] 堀, 古井，信学技報, SP2001-109, pp.43–48 (2001).

[4] C. Hori and S. Furui, Proc. EUROSPEECH2001, vol.III, pp.1771–1774, Aalborg (2001).

[5] http://www.cs.jhu.edu/~brill

[6] A.Waibel et al., Proc. HLT2001, pp.11–13, San Diego (2001).

[7] http://www.cis.upenn.edu/~treebank