

## 転移学習による抽出型要約の精度向上

西川 仁<sup>†1</sup> 牧野 俊朗<sup>†1</sup> 松尾 義博<sup>†1</sup>

異なる複数のドメインの参照要約が用意されているとき、テスト事例と異なるドメインの訓練事例も用いることによって抽出型要約の精度を向上させる方法を提案する。本稿では訓練事例のドメイン数に応じて特徴ベクトルを拡張する Augmented Space Method を用いて転移学習を実現する。学習時には構造学習を用いて学習を行う。本稿では要約の対象として6つの異なるドメインのコンタクトセンタ通話を取り上げ、それらの書き起こしを要約する。実験の結果、6ドメイン中4ドメインで提案手法が最良の結果を得た。

### Transfer Learning for Improving Quality of Extractive Summarization

HITOSHI NISHIKAWA,<sup>†1</sup> TOSHIRO MAKINO<sup>†1</sup>  
and YOSHIHIRO MATSUO<sup>†1</sup>

We propose a method to improve the quality of extractive summarization by making use of training examples whose domains are different from test examples. When references of multiple different domains are prepared, in this paper, we adopt Augmented Space Method that extends feature vectors depending on the number of domains to implement transfer learning. When learning we use a structured learning to learn parameters. As a target of summarization, we take up contact center dialogues in six domains and summarize their transcripts. As a result of our experiment, our proposed method achieved good results in four of six domains.

<sup>†1</sup> NTT サイバースペース研究所  
NTT Cyber Space Laboratories

#### 1. はじめに

単一、あるいは複数の文書からなる文書集合を計算機に要約させる際には、文書集合を文に分割し文集合を得て、与えられた要約長を超えないようにそれらの文集合の部分集合を選択し、選択された文を適切に結合することで要約を作ることが多い。この要約の方法を抽出型要約と呼ぶ<sup>10)</sup>。

抽出型要約を行う際には、与えられた文集合の部分集合に対して何らかの関数を定義し、その関数の出力値を、入力された部分集合の要約としての良し悪しを計るスコアとして用いることが多く、一般に、より高い値を得る文集合がより良い要約とされる。そのため、良い要約を得るためには、要約する文書の性質や、要約が用いられる場面等に応じて、必要とされる要約にその関数が高いスコアを与えるよう、その関数を適切に設計することが不可欠である。以下、その関数を目的関数と呼ぶ。

良い目的関数を得る一つの方法は、予め、要約の対象とする文書集合と、その文書集合に対する理想的な要約を何らかの方法で準備しておき、それらを用いてその関数を作ることである。ここでは、ある文書集合に対する理想的な要約を参照要約と呼び、要約対象の文書集合と参照要約の組を訓練事例と呼ぶ。この訓練事例が多数あれば、ある入力文書集合が与えられた際に出力すべき要約の傾向を学習することができ、この学習によって得られた関数を用いて新しく入力された文書集合を要約すれば、比較的良好な要約を得られる可能性が高い。

一方、上述したように、得たい要約の性質は、入力される文書集合の性質や、要約の用途によって大きく異なる。例えば、入力される文書集合は一つの新聞記事で、それを要約することを考える。このとき、記事が政治に関するものである場合と、経済に関するものである場合では、好まれる要約の性質が異なる可能性がある。要約の対象とする文書集合の性質に応じて適切な要約を作るため、入力される文書集合を適当な粒度で分類し、その分類に応じて、分類に対応した訓練事例から作った関数を用いることで、入力の性質に応じた要約を作ることができる可能性がある。ここでは、上述した、例えば政治や経済といった、入力される文書集合のある性質をまとめてドメインと呼ぶ。上の例では新聞記事には、政治ドメインに属する記事や、経済ドメインに属する記事があるということになる。

ドメインに応じて学習に必要な訓練事例を得ることができればそれぞれのドメインにおいて良い要約を得ることができると考えられるが、ドメイン毎に訓練事例を用意する必要がある。しかし、一般に訓練事例の用意には高いコストが必要となるため、それぞれのドメイ

ンにおいて必ずしも十分な数の訓練事例を得ることができるとは限らない。

そこで、本稿では、異なるドメインの訓練事例も利用して、要約の対象とするドメインの要約に用いる関数を得ることを考える。上の例を用いると、政治ドメインの新聞記事を要約するための関数を得ようとする際に、政治ドメインの訓練事例と共に経済ドメインの訓練事例も用いることで、より良い関数を得ようというものである。

本稿では、要約の対象としてコンタクトセンタ通話を取り上げる。これは、何らかの質問に対する回答を得るため、あるいは何らかの要求を達成するためにコンタクトセンタに電話をかけるもの（以下、話者 A とする）と、その電話を受けて話者 A に対応するオペレータ（以下、話者 B とする）による通話を書き起こした文書である。利用するコーパスの詳細は後述する。

以下、本稿の構成を述べる。2 節では本稿と深く関係する過去の研究を概説し、本稿の新規性を明らかにする。3 節では本稿で用いる要約モデルについて詳述する。4 節では要約モデルが必要とするパラメタの転移学習について述べる。5 節では要約の推論について述べる。6 節では提案手法の有効性を検証するため行った実験について述べ、7 節では本稿をまとめ、今後の研究の方向性についても述べる。

## 2. 関連研究

コンタクトセンタ通話の要約を提案している文献は既に存在している<sup>2),5),6)</sup>。Byrd らはコンタクトセンタ通話をヒューリスティックな規則を用いて要約する方法を提案している<sup>2)</sup>。Higashinaka らは、ドメインに特徴的な発話を発見する HMM を訓練し、発話列に対してラベリングを行うことで要約を行う手法を提案しており<sup>5)</sup>、また発話の重要度と HMM の出力を組み合わせて要約を行う手法も提案している<sup>6)</sup>。これらの文献では異なるドメインの訓練事例を陽に用いて要約の精度を高めるといことは行われていない。

要約の対象をコンタクトセンタ通話に限らず音声全般に広げると、更に多くの文献が存在している<sup>9),11)-13),15),16)</sup>。これらは全て音声を人手によって書き起こした結果、あるいは自動的に認識した結果を発話や文といった細かい単位に分割し、それらの単位に対して何らかの手法で重要度を付与し、重要度に応じてそれらの発話や文を選択することで要約を行っている。重要度を付与する方法として、コーパスから求めた tf-idf などの統計量<sup>9),11),15)</sup>、ロジスティック回帰を用いて訓練事例から計算された確率値<sup>12)</sup>、SVM によって得られた分離平面からの距離<sup>16)</sup> などがある。より正確に重要度を付与するため、参照要約の付与されている訓練事例と共に参照要約の付与されていない文書も用いる半教師あり学習を用いる方

法も提案されている<sup>16)</sup>。

本稿に最も近い先行研究として、電子メールのスレッドを要約するためにミーティングの音声の参照要約を利用することで精度向上を図るものがある<sup>13)</sup>。しかし、結果としては、転移学習を行わずに同一のドメインの訓練事例のみを用いて要約を行った場合が最良の結果を示しており、転移に成功しているとは言えない。これは、電子メールのスレッドの要約に対して、ミーティングの音声という大きく異なる性質を持つ訓練事例を転移させることが難しいためと考えられる。

音声の書き起こしを要約する際に、文そのものと、要約を構成する単語に対して別々にスコアを与える目的関数の有効性が示されている<sup>6),15)</sup>。本稿でもその形をした目的関数を用いるが、その場合、単語のスコアと文のスコアの要約全体への寄与度を決定するパラメタを何らかの方法で調整する必要がある。過去の文献ではこのパラメタを陽に調整していないか<sup>6)</sup>、あるいは人手で調整を行っている<sup>15)</sup>。このパラメタの調整の手間を避けるため、本稿では要約のパラメタ全体を構造学習として学習する。この方法を取ることににより、2つの要素の要約への寄与度を決定するパラメタは自動的に調整される。

要約のパラメタ推定を構造学習として行うものでは、MERT を用いるもの<sup>1)</sup>、Structured SVM を用いるものがある<sup>14)</sup>。これらはいずれも本稿とは異なる目的関数を用いており、また本稿では学習アルゴリズムとして Online Passive-Aggressive Algorithm<sup>3)</sup> を用いてパラメタを推定する。

ここで、先行研究と比較した際の本稿の新規性についてまとめておく。

- コンタクトセンタ対話のドメインの多様性に対応するため、転移学習を導入する。これによって単一のドメインからパラメタを学習した際よりもよい結果を得ることが可能となることを実験によって示した。
- 文と単語それぞれに関してスコアを与える要約モデルのパラメタを、構造学習として行う。そのため2つの項の寄与度を調整するパラメタの調整が不要となる。

## 3. 要約モデル

本稿では、要約の対象とする文の集合を  $\mathbf{x}$  と表記し、その部分集合であり与えられた要約長  $L$  を満たす要約を  $\mathbf{y}$  と表記する。また、要約の対象とする文の集合  $\mathbf{x}$  と事前に求めたパラメタ  $\mathbf{w}$  の下での、要約  $\mathbf{y}$  の実数値への写像  $\mathbf{f}_{\mathbf{x},\mathbf{w}} : \mathbf{y} \mapsto \mathbb{R}$  を要約にスコアを与える目的関数として用いる。このとき、要約問題は以下のように書ける。

表 1 文の特徴量  
Table 1 Features for sentences

特徴量	値域
正規化された文の位置	[0,1]
文が含むトークンの数	Integer
文が含む異なり語の数	Integer

表 2 単語の特徴量  
Table 2 Features for words

特徴量	値域
表記	0, 1
品詞	0, 1
入力文書集中での頻度	Integer
単語を含む文の入力文書中の数	Integer

$$\hat{\mathbf{y}} = \operatorname{argmax}_{\mathbf{y} \subseteq \mathbf{x}} \mathbf{f}_{\mathbf{x}, \mathbf{w}}(\mathbf{y}) \quad (1)$$

s.t.  $\text{length}(\mathbf{y}) \leq L$

上述したように、音声の書き起こしを要約する際に、文そのものと、要約を構成する単語に対して別々にスコアを与える目的関数の有効性が示されている<sup>6),15)</sup>。そのため、本稿でも要約を構成する文と単語に対して別々にスコアを与えるよう以下のように目的関数を設計する。

$$\mathbf{f}_{\mathbf{x}, \mathbf{w}}(\mathbf{y}) = \sum_{x_i \in \mathbf{y}} \mathbf{u}^\top \phi(x_i) + \sum_{z_j \in \mathbf{y}} \mathbf{v}^\top \psi(z_j) \quad (2)$$

ここで  $x_i$  は要約  $\mathbf{y}$  を構成する一つの文であり  $z_j$  は要約  $\mathbf{y}$  に含まれる一つの単語である。 $\mathbf{u} \in \mathbb{R}^{d_u}$  と  $\mathbf{v} \in \mathbb{R}^{d_v}$  はそれぞれパラメタベクトルである。 $\phi: x_i \mapsto \mathbb{R}^{d_u}$  と  $\psi: z_j \mapsto \mathbb{R}^{d_v}$  はそれぞれ文と単語の特徴ベクトルである。文と単語の特徴として、それぞれ表 1 と表 2 に示す特徴を用いる。

式 2 の 2 つの項のうち  $\sum_{x_i \in \mathbf{y}} \phi(x_i)$  と  $\sum_{z_j \in \mathbf{y}} \psi(z_j)$  はまとめて  $\Phi(\mathbf{x}, \mathbf{y})$  と表現するこ

とができ、同様にパラメタベクトル  $\mathbf{u}$  と  $\mathbf{v}$  をまとめて  $\mathbf{w}^\top = \langle \mathbf{u}^\top, \mathbf{v}^\top \rangle$  と表現することができる。そのため、目的関数は  $\mathbf{f}_{\mathbf{x}, \mathbf{w}}(\mathbf{y}) = \mathbf{w}^\top \Phi(\mathbf{x}, \mathbf{y})$  と表現することができる。

2 節で述べたように、本稿では  $\mathbf{w}$  を構造学習として直接最適化するため、2 つの項の寄与度を決定するパラメタを人手で調整する必要はない。

#### 4. 転移学習によるパラメタ推定

この節では、転移学習を用いて、異なる複数のドメインの訓練事例を利用し、要約のパラメタを推定する手法を提案する。まず転移学習の手法について述べ、その後具体的にパラメタを推定する際のアルゴリズムを示す。

##### 4.1 転移学習

まず、ここでは、実際に要約を作成したいドメインの  $N$  個の訓練事例  $\{(\mathbf{x}_j^t, \mathbf{y}_j^t)\}_{j=1}^N$  をターゲットドメインの訓練事例と呼ぶ。また、ターゲットドメインとは異なるドメインの  $M$  個の訓練事例  $\{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^M$  をソースドメインの訓練事例と呼ぶ。本稿では、ソースドメインもターゲットドメインもいずれもコンタクトセンタ通話を書き起こしたものであるが、ドメインによって通話のトピックが大きく異なる。コーパスについては 6 節で詳述する。

ターゲットドメインでの学習のためにソースドメインの訓練事例を援用する方法として、本稿では Augmented Space Method<sup>4)</sup> を用いる。この方法は学習法と独立に用いることができ、実装が容易という長所がある。品詞タグ付け問題等でその有効性を示している他<sup>4)</sup>、電子メールのスレッドの要約にも用いられている<sup>13)</sup>。

Augmented Space Method は、特徴ベクトル  $\Phi(\mathbf{x}, \mathbf{y})$  を以下のように拡張する。

$$\Phi^s(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}, \mathbf{y}), \Phi(\mathbf{x}, \mathbf{y}), \mathbf{0} \rangle$$

$$\Phi^t(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}, \mathbf{y}), \mathbf{0}, \Phi(\mathbf{x}, \mathbf{y}) \rangle$$

ここで、それぞれ  $\Phi^s$  はソースドメインの拡張された特徴ベクトル、 $\Phi^t$  はターゲットドメインの拡張された特徴ベクトルである。元の特徴ベクトルを  $n$  次元とし、ソースドメインとターゲットドメインが合わせて  $k$  ドメインあった場合、Augmented Space Method はソースドメインおよびターゲットドメインの特徴ベクトルを  $n \times (1+k)$  次元に拡張する。拡張された特徴ベクトルは、ドメイン間で共通である最初の  $n$  次元、そしてある 1 つのド

メインでしか利用されていない  $n$  次元, そして全てゼロとなっている残りの  $n \times (k-1)$  次元の要素からなる. なお, 上の例はターゲットドメインが1つ, ソースドメインが1つの場合だが, この手法は容易に2つ以上のソースドメインがある場合に拡張することができる. 特徴ベクトルに対応するパラメタベクトルも拡張され, 以降, 拡張された特徴ベクトルを  $\mathbf{w}'$  と表記する.

#### 4.2 構造学習

ここでは, 訓練事例よりパラメタベクトル  $\mathbf{w}$  を計算する方法について述べる. 本稿ではこのパラメタの学習を構造学習とみなして学習を行う. すなわち, 要約を構成する文や単語が要約に含まれる確率値等を個別に学習するのではなく, 文と単語の集合である要約を直接評価し学習を行う. 本稿では Online Passive-Aggressive Algorithm<sup>3)</sup> を用いてパラメタベクトルを訓練事例より推定する. パラメタベクトルの計算の際には1つずつ訓練事例が取り上げられ, 以下の更新式によってパラメタベクトル  $\mathbf{w}$  が更新される.

$$\mathbf{w}^{new} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{w} - \mathbf{w}^{old}\|^2 \quad (3)$$

$$\text{s.t. } \mathbf{w}^\top \Phi(\mathbf{x}_i, \mathbf{y}_i) - \mathbf{w}^\top \Phi(\mathbf{x}_i, \hat{\mathbf{y}}) \geq \ell(\hat{\mathbf{y}}; \mathbf{y}_i)$$

ここで,  $\mathbf{w}^{old}$  は更新前のパラメタベクトルであり,  $\mathbf{w}^{new}$  は更新後のパラメタベクトルである.  $\ell$  は損失関数である. これを解くと以下ようになる.

$$\mathbf{w}^{new} = \mathbf{w}^{old} + \lambda(\Phi(\mathbf{x}_i, \mathbf{y}_i) - \Phi(\mathbf{x}_i, \hat{\mathbf{y}})) \quad (4)$$

$$\lambda = \frac{\ell(\hat{\mathbf{y}}; \mathbf{y}_i) - s(\mathbf{x}_i, \mathbf{y}_i; \mathbf{w}^{old}) + s(\mathbf{x}_i, \hat{\mathbf{y}}; \mathbf{w}^{old})}{\|\Phi(\mathbf{x}_i, \mathbf{y}_i) - \Phi(\mathbf{x}_i, \hat{\mathbf{y}})\|^2} \quad (5)$$

本稿では, 損失関数に ROUGE<sup>8)</sup> を組み入れ, 以下のようにする.

$$\ell(\hat{\mathbf{y}}; \mathbf{y}_i) = 1 - \text{ROUGE}(\hat{\mathbf{y}}; \mathbf{y}_i)$$

損失関数に ROUGE を組み入れることで, ROUGE の値に沿ったパラメタを得ることができる. 本稿では ROUGE の中でも ROUGE-1 を用いた.

表 3 学習アルゴリズム  
Table 3 Learning algorithm

```

INPUT   $\{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^M, \{(\mathbf{x}_j^t, \mathbf{y}_j^t)\}_{j=M+1}^{M+N}$ 
INPUT   $\Phi'$ 
SET     $\mathbf{w}' = \mathbf{0}$ 
WHILE  parameter  $\mathbf{w}'$  isn't convergent
       $k \sim \text{random}(1, M+N)$ 
       $\hat{\mathbf{y}} = \underset{\mathbf{y} \subseteq \mathbf{x}_k}{\operatorname{argmax}} \mathbf{w}'^\top \Phi'(\mathbf{x}_k, \mathbf{y})$ 
       $\mathbf{w}' = \mathbf{w}' + \lambda(\Phi'(\mathbf{x}_k, \mathbf{y}_k) - \Phi'(\mathbf{x}_k, \hat{\mathbf{y}}))$ 
ENDIF
ENDWHILE
OUTPUT  $\mathbf{w}'$ 

```

表 4 推論アルゴリズム  
Table 4 Decoding algorithm

```

INPUT   $\mathbf{w}', \mathbf{x}_i, \Phi', L$ 
SET     $\mathbf{y} = \emptyset$ 
SET     $\mathbf{d} = \mathbf{x}_i$ 
WHILE   $\mathbf{d} \neq \emptyset$ 
       $\hat{\mathbf{y}} = \underset{\mathbf{y} \in \mathbf{d}}{\operatorname{argmax}} \frac{\mathbf{w}'^\top \Phi'(\mathbf{x}_i, \mathbf{y} \cup \hat{\mathbf{y}}) - \mathbf{w}'^\top \Phi'(\mathbf{x}_i, \mathbf{y})}{\text{length}(\hat{\mathbf{y}})}$ 
      IF   $\text{length}(\mathbf{y} \cup \hat{\mathbf{y}}) \leq L$  and
           $\mathbf{w}'^\top \Phi'(\mathbf{x}_i, \mathbf{y} \cup \hat{\mathbf{y}}) - \mathbf{w}'^\top \Phi'(\mathbf{x}_i, \mathbf{y}) \geq 0$  THEN
           $\mathbf{y} = \mathbf{y} \cup \hat{\mathbf{y}}$ 
        ENDIF
       $\mathbf{d} = \mathbf{d} \setminus \hat{\mathbf{y}}$ 
    ENDWHILE
       $\mathbf{y}^* = \underset{\mathbf{y} \in \mathbf{x}_i}{\operatorname{argmax}} \{ \mathbf{w}'^\top \Phi'(\mathbf{x}_i, \mathbf{y}) : \text{length}(\mathbf{y}) \leq L \}$ 
    OUTPUT  $\mathbf{y} = \underset{\mathbf{y} \in \{\mathbf{y}, \{\mathbf{y}^*\}\}}{\operatorname{argmax}} \mathbf{w}'^\top \Phi'(\mathbf{x}_i, \mathbf{y})$ 

```

具体的な学習アルゴリズムを表3に示す. 表3中のアルゴリズムでは, 煩雑さを避けるため, ソースドメインのパラメタベクトル  $\Phi^s$  もターゲットドメインのパラメタベクトル  $\Phi^t$  も  $\Phi'$  と表記した.

#### 5. 推論

推論アルゴリズムは, パラメタを学習する際と, 実際に要約を生成する際に用いられる. いずれの場合も同じものを使うことができる. アルゴリズムを表4に示す.

表4に示すアルゴリズムは Khuller らによるもので, このアルゴリズムによって最適解

表 5 通話のドメイン  
Table 5 Domain of dialogues

ドメイン	トピック
FIN	銀行や保険会社への問い合わせ
ISP	インターネットサービスプロバイダへの問い合わせ
LGU	地方自治体への問い合わせ
MO	通信販売会社への問い合わせ
PC	コンピュータメーカーへの問い合わせ
TEL	電話会社への問い合わせ

表 6 コーパスの統計量  
Table 6 Statistics of corpus

ドメイン	訓練事例				テスト事例			
	事例数	文数	文字数	平均要約率	事例数	文数	文字数	平均要約率
FIN	59	10377	110233	13.4%	60	8863	91946	16.3%
ISP	64	7062	98249	16.3%	59	9563	94355	15.6%
LGU	76	8865	89621	21.2%	56	7934	75948	18.4%
MO	70	9694	100680	17.4%	47	7305	58103	20.2%
PC	56	10088	105905	13.2%	44	11772	109917	10.0%
TEL	66	9774	98293	16.8%	41	8069	75620	13.6%

を得られる保証はないが、比較的良好な近似を得ることができる<sup>7)</sup>。

アルゴリズムは、現在の要約に新しく文を追加した際に増加するスコアをその文の長さで割った値、すなわち長さあたりの増加スコアが最大の文を要約に追加していくもので、全ての文を評価し終わった時点で要約候補を1つ得る。そのうち、入力文集合の中から、要約長を満たす文のうち最大のスコアを持つものを選び出し、先の要約候補と比較する。最終的にこの2つの要約候補のうち得点の高い方を要約として出力する。

## 6. 実験

### 6.1 コーパス

本稿ではコンタクトセンタ通話を書き起こしたものをを用いる。書き起こしは人手によって行われたもので、音声認識器の結果は用いない。これは音声認識器の誤りが結果に与える影響を除外するためである。本稿で用いるコンタクトセンタ通話は実際の通話ではなく、実際の通話を模擬して作成されたものである。それぞれの通話はコンタクトセンタに電話をかける話者 A とその電話を受け応対を行う話者 B の対話となっている。書き起こされた通話は既に発話に分割されている。一方の話者が何らかの発話を始めてから、もう一方の話者に発話が切り替わるまでが一つの発話となっている。本稿ではこの一つの発話を要約を構成する最小単位、すなわち文とする。

コーパス中の通話は6つのドメインに分かれている。それぞれのドメインのトピックを表5に示す。6つのドメインそれぞれに訓練事例が用意されているときに、それぞれのドメインの要約の精度を、他のドメインの訓練事例も用いて向上させることが本稿の目的である。

コーパスの統計量を表6に示す。事例数はそれぞれドメインにおいて訓練事例とテスト事例に含まれる事例の数、文数、文字数はそれぞれ事例全体に含まれる文の数、文字数である。平均要約率は、それぞれのドメインにおいて、入力文書の平均文字数を参照要約の長

さで割ったものである。訓練事例、テスト事例ともに参照要約の長さは250文字となっている。

### 6.2 実験設定

転移学習の先行研究に基づき、以下の4つの手法を比較する。

- (1) ソースドメインのみ6つのドメインのうち、要約を生成するドメインと同じドメインの訓練事例を除いて要約を生成する。例えばFINドメインの要約を生成する際には、他の5つのドメインを1つにまとめた訓練事例から学習したパラメタを用いて要約を生成する。
- (2) ターゲットドメインのみ6つのドメインそれぞれで、同じドメインの訓練事例のみを用いて学習を行い、その結果得たパラメタで要約を生成する。
- (3) 全ドメイン ソースドメイン、ターゲットドメインの区別なく訓練事例を1つにまとめ、それらを用いて学習を行い、得たパラメタから要約を生成する。
- (4) 転移学習 提案手法。上に述べた手法を用いてパラメタを学習し、それを用いて要約を生成する。

以降、上で述べた手法をそれぞれ手法(1)から手法(4)とする。

上で述べたように参照要約の文字数は250文字であるため、要約長も250文字とした。

### 6.3 評価

評価法は ROUGE-1<sup>8)</sup> を用いる。n-gram を構成する単語としては名詞、動詞、形容詞のみを用いる。

### 6.4 結果

結果を表7に示す。それぞれの行、列の数値は、対応するドメインのテスト事例を用いて、対応する手法で要約を行ったときの ROUGE 値である。それぞれのドメインで最も高

表 7 ドメイン, 手法別の ROUGE-1 値  
Table 7 ROUGE-1 result by domains and methods

ドメイン	手法 (1)	手法 (2)	手法 (3)	手法 (4)
FIN	0.437	0.557	0.532	<b>0.584</b>
ISP	0.465	0.497	0.505	<b>0.543</b>
LGU	0.466	0.528	<b>0.541</b>	0.500
MO	0.563	0.627	0.598	<b>0.635</b>
PC	0.215	<b>0.367</b>	0.307	<b>0.367</b>
TEL	0.421	0.416	<b>0.432</b>	0.418

い ROUGE 値の要素を太字にしてある。

6 ドメイン中, FIN, ISP, MO の 3 ドメインで, 提案手法が最も良い精度を得た。PC ドメインでは手法 (2) と同じ精度であった。LGU ドメインおよび TEL ドメインではいずれも手法 (3) に及ばなかった。

### 6.5 考 察

FIN ドメインでは提案手法が最も高い精度を達成し, ついで同ドメインの訓練事例を用いた手法 (2), ついで手法 (3), 手法 (1) となっている。この傾向は MO ドメインでも見られる。これら 2 つのドメインでは訓練事例のみで一定の精度が達成できるものの, 他のドメインの事例も利用することでより良い精度を達成することができている。

ISP ドメインでは提案手法が最も高い精度を達成している点では FIN ドメインおよび MO ドメインと変わらないが, 手法 (4) に続いて手法 (3), 手法 (2) と続いており, 同ドメインの訓練事例のみを用いた手法 (2) が手法 (3) に及ばない。ISP ドメインでの手法 (1) と手法 (2) を見ると, 他のドメインの手法 (1) と手法 (2) の精度の差と比べ差が大きい。このことから, ソースドメインの中に ISP ドメインのテスト事例に役立つ事例が多く含まれていることが示唆される。結果, 単純に訓練事例をまとめただけの手法 (3) が手法 (1), 手法 (2) を上回ったと考えられる。このような場合, 他のドメインの事例をうまく活用できることから, 表 7 が示すように転移学習は有効に働く。

LGU ドメインでは, 手法 (3) が最も良い性能を示しており, 次いで手法 (2), 手法 (4) となっている。他の 5 ドメイン全てにおいて手法 (4) は手法 (2) と同等か, あるいは上回っているが, LGU ドメインでは手法 (2) が手法 (4) を上回っている。一方, 手法 (1) の精度は高くないが, 手法 (3) の精度は高い。この結果は解釈が難しく, 今後の課題としたい。

PC ドメインでは, 同ドメインの訓練事例のみを用いた手法 (2) と手法 (4) の精度が等しく, この 2 手法を追って手法 (3), 手法 (1) という順となっている。これはこの PC ドメイ

ンは他のドメインと比べて大きく性質が異なっているため, 他のドメインの訓練事例がほとんど役に立たないことを示唆している。実際, PC ドメインにおける, 手法 (2) と比較したときの手法 (1) の精度は, 他のドメインに比べて著しく低い。手法 (2) と手法 (4) が同じ精度であるのも, 他のドメインの訓練事例が役に立たないため, 転移学習を行っても実際は同ドメインの訓練事例のみで訓練を行っているものと考えられる。

TEL ドメインでは, 手法 (2) が手法 (1) に及んでいない。すなわち同じドメインの訓練事例から推定したパラメタよりも, 他の 5 つのドメインの訓練事例をまとめて推定したパラメタを用いた方が良い結果が得られている。考えられるのは, TEL ドメインでは訓練事例とテスト事例でそれぞれの事例の性質が大きく異なっている可能性である。同一ドメインの事例のみを用いた手法 (2), 同一ドメインの事例を, 特徴ベクトルの拡張を通して「重く見る」手法 (4) の精度が低いことから, 他のドメインに TEL ドメインのテスト事例に大きく寄与する訓練事例が存在していることが示唆される。先行研究<sup>4)</sup>によれば, ソースドメインのみによる手法 (1) がターゲットドメインによる手法 (2) に上回る場合には, 転移学習, すなわち手法 (4) には効果がない。TEL ドメインの結果はその考察を裏付けるものと言える。

総じて, FIN や ISP, MO などの手法 (2) がよく働くドメインでは他のドメインの訓練事例を利用して精度の向上に成功している。他のドメインと比べ事例の性質が異なっているとみられる PC ドメインでは転移学習の効果がないが, 悪影響もない。TEL ドメインではそもそも同ドメインの訓練事例があまり役に立っておらず, そのため転移学習の結果も失敗している。

これらのことから, 同ドメインの訓練事例のみである程度働くドメインでは提案手法は有効であるが, そもそも与えられた訓練事例がテスト事例と異なる傾向を持つと考えられる場合などには転移学習の効果はあまりないといえる。また, 他のドメインと性質が大きく異なっているドメインに対しては他のドメインの訓練事例が役に立たず, 従って転移学習の効果がない。

## 7. おわりに

本稿では, 異なる複数のドメインの参照要約が用意されている状況で, テスト事例と異なるドメインの訓練事例も用いることで要約の品質を向上させる手法を提案した。学習時に特徴ベクトルを拡張して要約のパラメタを推定することで, 異なるドメインの参照要約を利用してより良いパラメタを得ることができることを示した。

また, 音声の書き起こしを対象にした要約のパラメタを構造学習によって推定することに

よって、これまで人手で調整していたパラメタを自動的に推定できることを示した。

今後の研究としては以下の方向性が考えられる。一つは、当然、全てのドメインにおいて手法 (1), (2), (3) を上回る学習法を考えることである。もう一つは、参照要約の付与されていない文書も利用して要約の精度を高める、半教師あり学習を利用する方向性である。コンタクトセンタ通話は膨大であるため、これらをうまく利用することができれば要約器の性能に大きな正の影響が得られるものと期待できる。

### 参 考 文 献

- 1) Aker, A., T. Cohn and R. Gaizauskas.: Multi-document summarization using A\* search and discriminative training, *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2010).
- 2) Byrd, R. J., M. S. Neff, W. Teiken, Y. Park, K. F. Cheng, S. C. Gates and K. Visweswariah: Semi-automated logging of contact center telephone calls, *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (2008).
- 3) Crammer, K., O. Dekel, J. Keshet, S. Shalev-Shwartz and Y. Singer: Online Passive-Aggressive Algorithms, *Journal of Machine Learning Research*, Vol. 7, pp. 551-585 (2006).
- 4) Daume III, H.: Frustratingly easy domain adaptation, *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics* (2006).
- 5) Higashinaka, R., Y. Minami, H. Nishikawa, K. Dohsaka, T. Meguro, S. Takahashi and G. Kikui.: Learning to model domain-specific utterance sequences for extractive summarization of contact center dialogues, *Proceedings of the 23rd International Conference on Computational Linguistics* (2010).
- 6) Higashinaka, R., Y. Minami, H. Nishikawa, K. Dohsaka, T. Meguro, S. Kobashikawa, H. Masataki, O. Yoshioka, S. Takahashi, and G. Kikui: Improving HMM-based Extractive Summarization for Multi-Domain Contact Center Dialogues, *Proceedings of IEEE Workshop on Spoken Language Technology* (2010).
- 7) Khuller, S., A. Moss and J. Naor.: The budgeted maximum coverage problem, *Information Processing Letters*, Vol. 70, No. 1, pp. 39-45 (1999).
- 8) Lin, C.-Y.: ROUGE: A package for automatic evaluation of summaries, *Proceedings of ACL Workshop on Text Summarization* (2004).
- 9) Lin, H., J. Bilmes and S. Xie: Graph-based Submodular Selection for Extractive Summarization, *Proceedings of the 11th Biannual IEEE Workshop on Automatic Speech Recognition and Understanding* (2009).
- 10) Mani, I.: *Automatic Summarization*, John Benjamins Pubs Co (2001).
- 11) Murray, G. and S. Renals: Towards Online Speech Summarization, *Proceedings of Interspeech* (2007).
- 12) Murray, G. and G. Carenini: Summarizing Spoken and Written Conversations. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2008).
- 13) Sandu, O., G. Carenini, G. Murray and R. Ng: Adaptation to Summarize Human Conversations, *Proceedings of ACL Workshop on Domain Adaptation in NLP* (2010).
- 14) Takamura, H. and M. Okumura: Learning to Generate Summary as Structured Output, *Proceedings of the ACM Conference on Information and Knowledge Management* (2010).
- 15) Xie, S., B. Favre, D. Hakkani-Tur and Y. Liu: Leveraging Sentence Weights in a Concept-based Optimization Framework for Extractive Meeting Summarization, *Proceedings of Interspeech* (2009).
- 16) Xie, S., H. Lin and Y. Liu: Semi-Supervised Extractive Speech Summarization via Co-Training Algorithm, *Proceedings of Interspeech* (2010).