圧縮型要約のオラクルに関する考察

平尾努　　西野正彬　　永田昌明
日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
{hirao.tsutomu,nishino.masaaki,nagata.masaaki}@lab.ntt.co.jp

1 はじめに

要約システムが出力可能かつある評価関数を最大化する要約であるオラクル要約は，要約システムのエラー分析や要約アルゴリズムの訓練データに利用できるなど，自動要約研究分野において重要な役割を担う．さらに，異なる要約モデルによって生成されたオラクル要約の評価スコアを比較することで，ある要約モデルがどの程度有用かを議論できるようにもなる．

これまで，自動要約研究の主流は文書中の文を抽出することで要約を生成する文抽出型要約 (*extractive summarization*)[Luhn 58] であったが，近年，文の文法性を保持するよう文書中の単語，あるいは文節を抽出する，つまり構文木の根付き部分木を抽出することで要約を生成する圧縮型要約 (*compressive summarization*)[富田 09，Berg-Kirkpatrick 11] の研究が盛んに行われるようになってきた．この手法は，文抽出と文圧縮を併用した要約生成法とみなすことができる．文を任意の長さに圧縮することによって，文抽出型要約よりも多くの文を原文書から抽出できることから，情報の網羅性が高い要約を生成できるという利点がある．しかし，文を圧縮したことにより可読性が損なわれたり，原文とは異なる意味の文が生成されるなど文章の品質が劣るという問題点がある．一方，文抽出型要約は文書中の文を抽出して要約を生成するため，少なくとも 1 文の意味，文法性は担保されるものの，予め定められた長さの文を抽出することしかできないため，要約長の制約が厳しい場合には抽出できる文の数が少なくなり，情報の網羅性が劣化するという問題がある．

本稿では，文抽出型要約，圧縮型要約手法のそれぞれによって生成されるオラクル要約，参照要約に対する ROUGE スコア [Lin 04] を最大化する要約を比較分析することで，圧縮型のオラクル要約の特徴を議論する．なお，双方の手法ともオラクル要約の生成は NP 困難な組合せ最適化問題であり，最適解を得るための多項式時間アルゴリズムは知られていない．よって，最悪の場合ではそれぞれ，文の数，単語数に対して指数オーダの計算量が必要となる．そこで，本稿ではオラクル要約生成を整数計画問題 (Integer Linear Programming Problem) して定式化し，分枝限定法を活用した ILP ソルバを用いて実用上は問題ない程度の時間でオラクル要約を生成する手法を提案する．

TSC-3 (Text Summarization Challenge) コーパス [Hirao 04] を用いて，参照要約に対するオラクル要約を生成した結果，圧縮型のオラクル要約の平均 ROUGE スコアは文抽出型のオラクル要約のそれと比較して 0.1 ポイント以上高いことがわかった．しかし，圧縮型のオラクル要約は ROUGE に最適化するあまり，小さな (文節数が少ない) 根付き部分木を多数抽出する傾向にあり，その結果，ROUGE の改善には寄与するものの，文章としての品質が低いことがわかった．さらに，こうした圧縮型要約の問題点を改善する一つの方法として，抽出する文節数に制約を設けることが有効であることもわかった．

2 オラクル要約生成の整数計画問題としての定式化

2.1 オラクル要約の定義

オラクル要約生成の整数計画問題としての定式化を与える前に，ROUGE の定義を説明し，オラクル要約を定義する．

R を参照要約，$\mathcal{R}$ を参照要約に出現する N グラムの多重集合，S をシステム要約，$\mathcal{S}$ をシステム要約に出現する N グラムの多重集合，$U(\cdot)$ は多重集合を集合へと変換する関数とし，$C(t_n,\mathcal{R})$，$C(t_n,\mathcal{S})$ を N グラム t_n の多重集合 $\mathcal{R}$，$\mathcal{S}$ における頻度とすると，R と S との間の ROUGE_n スコアは，以下の式で定義される [Lin 04]．

$$\text{ROUGE}_n(R,S) = \frac{\sum_{t_n \in U(\mathcal{R})} \min\{C(t_n,\mathcal{R}), C(t_n,\mathcal{S})\}}{\sum_{t_n \in U(\mathcal{R})} C(t_n,\mathcal{R})}. \quad (1)$$

本稿でのオラクル要約とは ROUGE を最大化するシステム要約なので，オラクル要約 O を以下の式で定義する．

$$O = \arg\max_{S \subseteq D} \text{ROUGE}_n(R,S) \quad s.t. \quad \ell(S) \leq L_R \quad (2)$$

$\ell(\cdot)$ は，要約の長さを返す関数，D は文抽出型のオラクル要約の場合，文書 (セット) 中に含まれる文の集合，圧縮型のオラクル要約の場合，文書中に含まれる文集合から得られる可能なすべての根付き部分木の集合，L_R は参照要約の長さである．これは NP 困難な組合せ最適化問題である．

2.2 文抽出型オラクル要約生成の整数計画問題としての定式化

式 (1) より，ROUGE の分母は，参照要約に含まれる N グラムの頻度の和であることから定数である．よって，式 (2) より，ある長さの制約のもとで ROUGE を最大化することは，式 (1) の分子を最大化することに等しい．いま，$|U(\mathcal{R})| = M$，として集合中の j 番目の N グラムに対する $\min\{C(t_n, \mathcal{R}), C(t_n, \mathcal{S})\}$ の値を z_j とおくと，目的関数は以下の式となる．

$$\underset{\boldsymbol{z}}{\text{maximize}} \quad \sum_{j=1}^{M} z_j \tag{3}$$

また，制約条件は以下の式となる．

$$s.t. \qquad \sum_{i=1}^{\mathcal{N}} w_i x_i \leq L_R \tag{4}$$

$$\forall j \quad \textstyle\sum_i \text{count}_{i,j} x_i \geq z_j \tag{5}$$

$$\forall j \quad \text{refcount}_j \geq z_j \tag{6}$$

$$\forall i : x_i \in \{0, 1\} \tag{7}$$

$$\forall i : z_j \in \mathbb{Z}_+, \tag{8}$$

いま，総文数を $\mathcal{N}$ とし，w_i は i 番目の文の長さとする．x_i は，i 番目の文をオラクル要約として選択するか否かを決定する 0/1 変数であり，式 (4) は，オラクル要約の長さが L_R 以下であることを保証する．

z_j は j 番目の N グラムの参照要約における頻度とオラクル要約における頻度の小さい方の値でなければならない．いま，i 番目の文における j 番目の N グラムの頻度を $\text{count}_{i,j}$ とすると文集合全体での最大頻度の値は $\sum_i \text{count}_{i,j} x_i$ となる．一方，j 番目の N グラムの参照要約における頻度を refcount_j とすると，z_j が $\sum_i \text{count}_{i,j} x_i$，$\text{refcount}_j$ の双方より小さいという制約を導入すれば，z_j は，$\min\{C(t_n, \mathcal{R}), C(t_n, \mathcal{S})\}$ と等しくなる．

なお，我々は以前にある参照要約に対し，文抽出型のオラクル要約をすべて列挙することを目的としたアルゴリズムを提案した [平尾 14]．本稿の主たる目的はオラクル要約の列挙ではなく，文抽出型，圧縮型のそれぞれのオラクル要約間の ROUGE スコアの比較である．そこで，オラクル要約生成を列挙のいらない整数計画問題として定式化し，ILP ソルバを利用してオラクル要約を得た．

2.3 圧縮型オラクル要約生成の整数計画問題としての定式化

本稿では，要約対象文書の言語を日本語とし，文の構文構造は文節間の係りうけにより表されているとする．よって，本稿での圧縮型のオラクル要約とは，文節抽出によって生成されたものをさす．

いま，文書中の文節，(単語)N グラムにそれぞれ何番目の文か，その文の何番目の文節，N グラムかをあらわすインデックスが割り当てられているとする．圧縮型オラクル要約の生成問題は文抽出型オラクル要約の生成問題と同じく整数計画問題として定式化でき，その目的関数は文抽出型の場合 (式 (3)) と同じである．j 番目の N グラムの頻度 $z_j (\in \mathbb{Z}_+)$ に対する制約は，以下のとおりとなる．

$$\begin{aligned} \forall j \quad & \sum_{\tau \in \mathcal{T}_j} n_{<\tau>} \geq z_j \\ \forall j \quad & \text{refcount}_j \geq z_j \end{aligned} \tag{9}$$

z_j の値が参照要約における頻度よりも小さいという制約は文抽出型の場合と同じである．いま，文書中で j 番目の N グラムが出現する文番号，文節番号をあらわすインデックスのタプル集合を $\mathcal{T}_j$ とする．そして，i 番目の文の k 番目の N グラムをオラクル要約に含めるか否かをあらわす 0/1 変数を $n_{i,k}$ とすると，文集合全体でとることのできる j 番目の N グラムの最大頻度は $\sum_{\tau \in \mathcal{T}_j} n_{<\tau>}$ となるので，z_j がそれよりも小さくなるという制約を導入する．

さらに，抽出の最小単位が文節であることから，i 番目文の k 番目の N グラムをオラクル要約に含める場合には，それを含む K 個の文節もオラクルに含めなければならない．そこで以下の制約を導入する．

$$\forall i, \ \forall k \quad b_{i,k(1)} \geq n_{i,k}, \ldots, b_{i_k(K)} \geq n_{i,k} \tag{10}$$

なお，$b_{i,k}$ は i 番目の k 番目の文節をオラクル要約に含めるか否かをあら 0/1 変数である．

さらに，ある文から抽出した文節が根付き部分木を構成するように以下の制約を導入する．

$$\forall i, \ \forall k \quad b_{i,k} \leq b_{i,\text{parent}(k)} \tag{11}$$

$\text{parent}(j)$ は j 番目の文節の係り先文節のインデックスをあらわす．

さらに，$w_{i,k}$ を i 番目の文の k 番目の文節の長さとして，要約長の制約を以下の式であらわす．

$$\sum_i \sum_k w_{i,k} b_{i,k} \leq L_R \tag{12}$$

よって，最終的に式 (3) を目的関数，式 (9) から (12) を制約条件とした整数計画問題を解くことで圧縮型のオラクル要約を得ることができる．

3 検証実験

3.1 実験設定

実験には TSC (Text Summarization Challenge) 3 のコーパス [Hirao 04] を利用した．TSC-3 コーパスは 30 文書セット (トピック) から構成される．各セットはある特定のニュースに関する新聞記事を 10 記事

表 1: オラクル要約と TSC-3 のベストシステムの ROUGE スコア

	Short		Long	
	R-1	R-2	R-1	R-2
文抽出型	.726	.544	.770	.596
圧縮型	.870	.672	.896	.702
F0307	.492	.307	.542	.343

程度含み，その総文字数に対し，約 6%(以下，short)，12%(以下，long) の 2 種の長さの参照要約が与えられる．1 トピックあたりの平均文数は約 120 文，平均文字数は 6564 文字である．

単語分割，品詞タグ付けには MeCab を用い，分節係りうけ解析には CaboCha を用いた．整数計画問題の目的関数の計算対象とする N グラムは，ユニグラム ($ROUGE_1$) の場合とバイグラム ($ROUGE_2$) の双方を試した．ユニグラムは名詞，動詞のみを対象とし，バイグラムはすべての単語を対象とした．また，整数計画問題は，CPLEX(12.5.1.0) を用いて解いた．

3.2 オラクル要約の ROUGE スコア

表 1 に文抽出型，圧縮型のオラクル要約と TSC-3 で最良の結果を得たシステム (ID:F0307)[1] の $ROUGE_1$，$ROUGE_2$ スコアを示す．

表 1 より，F0307 とオラクル要約の ROUGE スコアを比較すると，明らかにオラクル要約の ROUGE スコアが高い．圧縮型オラクルとの差は 0.4 ポイント近く，文抽出型オラクルでもその差は 0.3 ポイント近くもある．この結果は，現状の文抽出型要約システム [2] にもまだまだ改良の余地が残っていることを示唆している．

文抽出型と圧縮型の ROUGE スコアを比較すると，圧縮型の ROUGE スコアは文抽出型のそれよりも 0.1 ポイント以上高い．これは，文圧縮を用いることで単なる文抽出よりも多くの「文」(根付き部分木) を抽出できるようになり，ROUGE 向上に寄与する N グラムを抽出できるようになったからだろう．

なお，係り受けは無視して単語抽出によりオラクル要約を生成したところ [3]，short の $ROUGE_1$ は 0.948，$ROUGE_2$ は 0.749，long の $ROUGE_1$ は 0.945，$ROUGE_2$ は 0.764 であった．参照要約には原文書にはない N グラムが存在するため，そもそもオラクル要約が理論上の ROUGE スコアの上限値である 1 をとることはない．しかし，圧縮型のオラクル要約の ROUGE スコアが係り受け制約のない単語抽出オラクルのそれと近い値であること，文抽出型オラクル要約の ROUGE スコアとの間に大きな差があることを考えると，圧縮型要約が ROUGE という観点から文抽出型要約よりも有望な手法であることは間違いない．

[1] このシステムは，クエリ指向の文抽出型要約システムである．

[2] F0307 は約 10 年前のシステムであるが，DUC においても約 10 年前のシステムと現状のシステムの差が 0.1 ポイント未満であることを考えると，現状のシステムがこの ROUGE スコアを 0.2 ポイント以上改善しているとは考え難い．

[3] 紙面の都合上詳細は割愛するが，これも整数計画問題として定式化できる．

表 2: 要約あたりの平均文数

	Short		Long	
	Unigram	Bigram	Unigram	Bigram
文抽出型	8.4	7.2	15.3	13.7
圧縮型	16.7	13.9	27.9	23.4
F0307	6.2		11.2	
参照要約	7.1		13	

表 3: 2 文節以下の文の割合

	Short		Long	
	Unigram	Bigram	Unigram	Bigram
圧縮型	.325	.248	.249	.161
参照要約	0.009		.008	
コーパス	.071			

3.3 圧縮型オラクル要約の問題点

圧縮型オラクル要約が非常に高い ROUGE スコアを持つことがわかったが，要約として必要とされる文章としての品質を保っているかどうかを要約に含まれる文の数，1 文あたりの長さ (文節数) という観点から考察する [4]．

表 2 にオラクル要約，F0307，参照要約の 1 要約あたりの平均文数 (圧縮型のオラクル要約の場合は根付き部分木) を示す．この結果より，抽出型オラクル要約に含まれる文の数は参照要約に含まれる文の数とほぼ同等であるが，圧縮型のオラクル要約の文の数は明らかに他の手法よりも多い．参照要約と比較すると約 2 倍程度の文数である．どの要約も上限として与えられる文字数制限は同じなので，この結果は圧縮型オラクル中の 1 文が他の手法よりも短いことを示している．

次に，圧縮型のオラクル要約，TSC-3 コーパスの記事全体，参照要約のそれぞれについて文中の文節数を調べ，2 文節以下の文が全体に占める割合と 1 文あたりの文節数の最頻値を調べた．その結果を表 3，4 にそれぞれ示す．

表 3 より，2 文節以下の文は，参照要約にはほぼ存在せず，TSC-3 コーパス全体でも 10%に満たない程度の数しか存在しない．しかし，圧縮型のオラクル要約ではそれが約 25%程度も存在している．また，表 4 より，1 文あたりの文節数の最頻値をみると，圧縮型オラクル要約の場合，それが 2 文節 ($ROUGE_2$ のみ 4 文節) である．コーパス全体での 1 文あたりの文節数の最頻値は 7，それの参照要約での最頻値は short で 9，long で 11 なので，これは明らかに小さな値である．2 文節以下の文の実例を以下に示す．

> 「中古市場が出なくなる。」,「東京地裁であった。」,「スハルト氏はなっていた。」,「年俸は推定されている。」,「ＹＳ１１はなった。」,「これにより、している。」

このように一般的に 2 文節以下の文は ROUGE スコアに寄与する N グラムを含んでいたとしても文章としての品質が低く，その意味を汲み取れるようなものではないものがほとんどである．

[4] もちろん，複数名の人間が文を読み判定することが望ましいが，本稿では擬似的に文の数と文節の数でこれを判定する．

表 4: 文節数の最頻値

	Short		Long	
	Unigram	Bigram	Unigram	Bigram
圧縮型	2	2	2	4
参照要約	9		11	
コーパス	7			

文抽出型オラクル要約には，文書中の文をそのまま抽出するため，当然，このように品質の低い文が要約に含まれることはない．よって，文抽出型のオラクル要約を ROUGE スコア最大の要約とすることは理にかなっているが，圧縮型の場合，先に述べたように文章の品質を犠牲にして簡単に ROUGE スコアを向上できるので，単に ROUGE 最大の要約をオラクルとすることには問題がある．ROUGE スコアの向上に最適化するあまり，文章としての品質が犠牲になってしまっていることを考えれば，人間が許容可能な文章の品質を保った上で ROUGE を最大化する要約をオラクルとすべきだろう．なお，圧縮型要約システムを構築する際にもこの知見は考慮すべきである．

3.4 圧縮型オラクル要約の改善

圧縮型のオラクル要約に短い文が多く含まれる理由の一つとして，抽出する文節が根付き部分木構成することだけを制約としていることがあげられる．式 (11) の制約では，係り受けの根ノードであればたとえ 1 文節であっても根付き部分木であるため制約を満たしてしまう．先に示したとおり，2 文節以下の文は文として体をなしていない可能性が高い．

これを抑える簡単な方法の一つとして，抽出する文節数に制約を設けることがあげられる．たとえば，ある文節数以下の根付き部分木は抽出できないようにすればよい．以下にその制約を示す．

$$\frac{\sum_{k=1}^{J} b_{i,k}}{\mathcal{K}} \geq a_{i,1} \quad (13)$$

$$1 - \frac{\sum_{k=1}^{J} b_{i,k}}{J} \geq a_{i,2} \quad (14)$$

$$a_{i,1} + a_{i,2} = 1 \quad (15)$$

J は i 番目の文の文節数，$a_{i,1}$，$a_{i,2}$ は 0/1 のバイナリ変数であり，式 (11) は，$a_{i,1} = 1$ のとき，i 番目の文から抽出する根付き部分木の文節数が $\mathcal{K}$ 以上であることを保証する．式 (12) は，$a_{i,2} = 1$ のとき，i 番目の文からは根付き部分木を抽出しないことを保証する．式 (13) によって，$a_{i,1}$，$a_{i,2}$ のどちらか一方が 1 をとることで，根付き部分木を抽出する際には必ずその文節数が $\mathcal{K}$ 以上であることを保証する．

コーパス中の文節数の最頻値が 7 であることから，$\mathcal{K} = 7$ とした場合の圧縮型オラクル要約の ROUGE スコア，文数を表 5 に示す．表 1，2 と比較すると，ROUGE スコアは 0.04 ポイント程度の低下しているが，文数は参照要約のそれに近い値となった．また，文節数の最頻値はすべての場合で 7 であった．

表 5: 文節数の制約付き圧縮型オラクル要約の ROUGE スコアと文数

	Short		Long	
	Unigram	Bigram	Unigram	Bigram
ROUGE	.820	.630	.852	.664
文数	9.30	8.90	17.5	16.9

実際にオラクル要約を読むと文節数制約がない場合と比較して適度な長さの圧縮文となっており文章の品質は明らかに向上していた．文章の品質を保ちつつ最大の ROUGE スコアはおよそ表 5 程度の値であろう．しかし，文節数制約により，どのような圧縮文でも 7 文節以上を抽出せねばならないため，文頭文節として「よると、」,「ことで、」,「ため、」などが数合わせで抽出されてしまい，可読性に悪影響をおよぼす場合もあった．こうした問題を解決するためには，文献 [Yoshikawa 12] のように，述語項構造解析の結果から述語とのその必須格を圧縮文に残すなど言語知識を有効活用すべきだろう．

4 おわりに

本稿では，文抽出型要約，圧縮型要約のオラクル要約生成の整数計画問題としての定式化を提案した．TSC-3 コーパスに対し双方の手法によるオラクル要約を生成し，ROUGE スコアを調べたところ，圧縮型要約は 0.1 ポイント以上その値が高いことがわかった．しかし，ROUGE に最適化するあまり短い文を抽出する傾向にあり，文章の品質が低いことも明らかとなった．この問題を解決するため，抽出する文節数に制約を設けたところ，ROUGE スコアはわずかに低下するものの，文章の品質は大きく改善できることを示した．

参考文献

[Berg-Kirkpatrick 11] Berg-Kirkpatrick, T., Gillick, D., and Klein, D.: Jointly Learning to Extract and Compress, in *Proc. of the 49th ACL*, pp. 481–490 (2011)

[Hirao 04] Hirao, T., Fukusima, T., Okumura, M., Nobata, C., and Nanba, H.: Corpus and Evaluation Measures for Multiple Document Summarization with Multiple Sources, in *Proc. of the 20th COLING*, pp. 535–541 (2004)

[Lin 04] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, in *Proc. of Workshop on Text Summarization Branches Out*, pp. 74–81 (2004)

[Luhn 58] Luhn, H. P.: The Automatic Creation of Literature Abstracts, *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159–165 (1958)

[Yoshikawa 12] Yoshikawa, K., Iida, R., Hirao, T., and Okumura, M.: Sentence Compression with Semantic Role Constraints, in *Proc. of the 50th ACL*, pp. 349–353 (2012)

[富田 09] 富田紘平, 高村大也, 奥村学：重要文抽出と文圧縮を組み合わせた新たな抽出的要約手法, 情報処理学会研究報告 (SIG-NL-189), pp. 13–20 (2009)

[平尾 14] 平尾努, 西野正彬, 鈴木潤, 永田昌明：オラクル要約の列挙, 言語処理学会年次大会, pp. 650–653 (2014)