

# 強化学習によるテキスト自動要約手法の提案

梁 成基<sup>†</sup>阿辺川 武<sup>‡</sup>東京大学 情報理工学系研究科<sup>†</sup>国立情報学研究所<sup>‡</sup>

{sryang, abekawa}@nii.ac.jp

## 1 はじめに

膨大なテキスト集合からその要約を自動的に生成する問題は、自然言語処理の分野と情報検索の分野において、古くから研究されている話題である [1].

近年では主に抽出的な手法、特に文を抽出単位とした手法が広く研究されており、文に対してスコア関数を定義し指定の制限長に収まる最大のスコアを持つ文集合を選択する問題として定式化されるのが典型的である。MMR を用いた貪欲的な手法は、代表的な抽出的自動要約手法である [2]. しかし、局所的には高いスコアを持つ文を貪欲的に一つ選択したとしても、要約全体として良い要約である保証は無く実際は大域的に最適化されなければならない。ナップザック問題として定式化した上で整数計画問題を最適化したり [5], 最大被覆問題として定式化して [7], 大域的な最適化問題にアプローチしているのが近年の傾向である。また文圧縮手法は、文に対する語の抽出的要約であると解釈でき、文抽出同様に多くの研究がなされている [3, 4]. 文圧縮においても文法的に妥当な文を生成しなければならないという制限があり、同様に大域的な最適化を行う必要がある。しかし整数計画問題は NP-hard であり、文圧縮まで考慮すると膨大な計算時間を必要とする問題点がある。

本稿では、自動要約問題を強化学習問題 [6] として定式化し要約を得る手法を提案する。今の要約の状態においてどの文を抽出すれば良いのか、またどのように文圧縮すればいいのかという、一つの人間の意思決定のモデルを素直に反映した枠組みで自動要約を試みる。文抽出問題のみを考えた問題を強化学習として定式化し、次にそれに既存の文圧縮手法を一つの強化学習における行動だと考え追加することによって、文圧縮についても考慮することが出来る事を見る。実験を行い、強化学習問題として解くことによって要約が得られている事を確かめ、文圧縮を考慮する事による効

果が表れている事を確かめた。

## 2 文抽出問題の定式化

本節では、我々の提案手法について述べる。以下、文抽出のみを考慮する要約問題であるとする。

与えられた文集合を  $X = \{x_1, x_2, \dots, x_N\}$  とする。文書集合  $D = \{X_1, X_2, \dots, X_M\}$  が与えられたならば、 $X = \bigcup_i X_i = \{x_1, x_2, \dots, x_N\}$  として考える。要約とは、部分集合  $S \subset X$  の事であり、 $L(S)$  は要約  $S$  の要約長を表す。  $K$  は要約長の上限である。

文抽出問題とは、与えられたスコア関数  $score(S)$ , 要約長の上限  $K$ , 要約長の許容範囲を定める定数  $C$  に対して、

$$S^* = \arg \max_{S \subset X} score(S) \quad (1)$$

$$\text{s.t. } K - C \leq L(S) \leq K \quad (2)$$

なる要約  $S^*$  を求める問題とする。

### 2.1 強化学習

強化学習 (Reinforcement Learning) は、ある与えられた環境に対して試行錯誤を通じて環境に自ら適応する学習方式である。学習者であるエージェントは (1) 現在の状態  $s_t$  を観測し、(2) その状態に対して一つ行動  $a_t$  を選択・実行し、(3) 報酬と呼ばれるスカラー量  $r_t$  と次の状態  $s_{t+1}$  を環境から観測する、というステップを継続する。学習の目標は状態に対して将来に得られるであろう報酬の和を最大化するような行動を決定出来るような方策  $\pi$  を求めることである。教師あり学習とは違い、真に取るべきであった行動の代わりにあるスカラー量を指標として与えるだけであり、これが唯一の情報である所がこの問題の特徴である。以下の節では、強化学習における状態、行動、報酬のモデルをどのように定義するかについて述べる。

## 2.2 状態

状態  $s$  は現在の要約に含まれる文集合を保持しており、任意の要約  $S$  の状態を表現する。状態集合全体  $\mathcal{S}$  は、可能な要約全体である。

$$s_t = \begin{pmatrix} S_t \\ A_t \\ f_t \end{pmatrix} \quad (3)$$

ただし、 $S_t$  は要約、 $A_t$  は状態  $s_t$  に至るまでにとった行動の集合、終了状態を表すビット  $f_t \in \{0, 1\}$  とする。これにより終了状態集合は  $\mathcal{S}_G = \{s | s \in \mathcal{S}, f = 1\}$  とする。初期状態はどの文も追加していない空の状態であり、 $s_0 = (\emptyset, \emptyset, 0)^T$  である。

## 2.3 行動

行動  $a$  は、途中過程の要約  $S$  に対して何らかの操作を加える操作である。文抽出の問題であると考えられる場合、行動集合全体  $\mathcal{A}$  は以下のような  $N + 1$  要素からなる集合と考える事が出来る。

$$\mathcal{A} = \{\text{insert}_1, \text{insert}_2, \dots, \text{insert}_N, \text{finish}\} \quad (4)$$

それぞれについて以下の様な決定的な遷移を与える。

$$\begin{pmatrix} S_t \\ A_t \\ 0 \end{pmatrix} \xrightarrow{\text{insert}_i} \begin{pmatrix} S_t \cup \{x_i\} \\ A_t \cup \{\text{insert}_i\} \\ 0 \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} S_t \\ A_t \\ 0 \end{pmatrix} \xrightarrow{\text{finish}} \begin{pmatrix} S_t \\ A_t \cup \{\text{finish}\} \\ 1 \end{pmatrix} \quad (6)$$

また、状態  $s_t$  においてとることが出来る行動集合  $A(s_t) (\subset \mathcal{A})$  を以下の様に定義する。

$$A(s_t) = \begin{cases} \{a | a \in \mathcal{A}, a \notin A_t\} & (L(S_t) \leq K) \\ \{\text{finish}\} & (K < L(S_t)) \end{cases} \quad (7)$$

指定された要約長を超えていなければ今までとっていない行動集合が可能な行動集合であり (finish を含む)、要約長を超えていれば強制的に終了することになる。

## 2.4 報酬

環境は、以下の様にエージェントに報酬  $r_t$  を与える。次の状態  $s_{t+1}$  が終了状態集合  $\mathcal{S}_G$  に含まれてい

る場合、つまり行動  $a_t = \text{finish}$  を実行した場合は、

$$r_t = \begin{cases} \text{score}(S_t) & (K - C \leq L(S_t) \leq K) \\ -R_{\text{penalty}} & (\text{otherwise}) \end{cases} \quad (8)$$

として報酬を与える ( $R_{\text{penalty}} > 0$ )。設定した要約長の許容範囲に含まれていればスコアを、そうでなければ罰を与えている。finish 以外であれば、 $r_t = 0$  とする。前項の可能行動集合  $A(s_t)$  の定義と組み合わせれば、要約長を超えている要約を生成してしまったならば強制的に終了してマイナスの報酬、つまり罰を受けることになり、うまく終了すればその要約のスコアが得られる事になる。

## 2.5 まとめ

前節までにおいて我々が見てきたモデルを要約問題における環境の定義とし、この環境に対して完全な状態価値関数を得、最適な方策を得たとすると、初期状態  $s_0$  から開始して最も報酬が得られる行動列とは、最もスコアが高い要約を持つ状態に至る行動列となる。

よってこの問題を強化学習問題として方策を求める事で、解が得られる。ただし、任意のあり得る要約を考えているため状態空間は爆発的に大きくなっており、強化学習問題としての定式化は出来たととしても、実時間内に解を求める事が出来るかは別問題である。状態空間をうまく素性空間にうつし、ある程度の近似解が求まるように出来ることは期待出来る。このことは4章の実験の結果で示す。

## 3 文圧縮手法を考慮した定式化

前章では、文抽出手法をどのように強化学習問題としたかを見た。前章までの定式化に、文圧縮手法を考慮に加える方法の一つの例を与える。

既存の文圧縮器が既に完成しているとして、ある文  $x$  に対して  $\text{comp}(x, \mu)$  として文圧縮できると仮定する。 $\mu$  は圧縮時に用いるパラメータとする。この例では、状態と行動に関して変更を必要とする。

### 3.1 状態

状態  $s$  は、前章の定義 (3) に一つ新しいビットを導入する。

$$s_t = \begin{pmatrix} S_t \\ A_t \\ c_t \\ f_t \end{pmatrix} \quad (9)$$

ただし,  $c_t \in \{0, 1\}$  とし, 初期状態は  $s_0 = (\emptyset, \emptyset, 1, 0)^T$  とする.

### 3.2 行動

(4) で定義した  $\mathcal{A}$  に対して, 新しい行動  $\text{comp}_\mu$  を導入する.

$$\begin{pmatrix} s_t \\ S_t \\ A_t \\ 0 \\ 0 \end{pmatrix} \xrightarrow{a_t, \text{comp}_\mu} \begin{pmatrix} s_{t+1} \\ (S_t \setminus \{x'\}) \cup \{\text{comp}(x', \mu)\} \\ A_t \cup \{\text{comp}_\mu\} \\ 1 \\ 0 \end{pmatrix} \quad (10)$$

ただし,  $x'$  は  $S_t$  の最後の要素とする.  $\text{comp}_\mu$  により, 抽出された最後の文をパラメータ  $\mu$  で圧縮した状態に遷移する. さらに, 新しく抽出された場合に圧縮出来るように  $\text{insert}_i$  と  $\text{finish}$  を変更する.

$$\begin{pmatrix} s_t \\ S_t \\ A_t \\ c_t \\ 0 \end{pmatrix} \xrightarrow{a_t, \text{insert}_i} \begin{pmatrix} s_{t+1} \\ S_t \cup \{x_i\} \\ A_t \cup \{\text{insert}_i\} \\ 0 \\ 0 \end{pmatrix} \quad (11)$$

$$\begin{pmatrix} s_t \\ S_t \\ A_t \\ c_t \\ 0 \end{pmatrix} \xrightarrow{a_t, \text{finish}} \begin{pmatrix} s_{t+1} \\ S_t \\ A_t \cup \{\text{finish}\} \\ c_t \\ 1 \end{pmatrix} \quad (12)$$

上記の  $\text{comp}_\mu$  定義から  $c_t = 0$  の時のみ圧縮可能なので,  $A(s_t)$  は以下の様に変更される.

$$A(s_t) = \begin{cases} \{a | a \in \mathcal{A}(\text{insert}), a \notin A_t\} \cup \{\text{finish}\} \cup \mathcal{A}(\text{comp}) & (L(S_t) \leq K, c_t = 0) \\ \{a | a \in \mathcal{A}(\text{insert}), a \notin A_t\} \cup \{\text{finish}\} & (L(S_t) \leq K, c_t = 1) \\ \mathcal{A}(\text{comp}) \cup \{\text{finish}\} & (K < L(S_t), c_t = 0) \\ \{\text{finish}\} & (K < L(S_t), c_t = 1) \end{cases} \quad (13)$$

ただしここで,  $\mathcal{A}(\text{insert}) = \bigcup_i \{\text{insert}_i\}$ ,  $\mathcal{A}(\text{comp}) = \bigcup_\mu \{\text{comp}_\mu\}$  と置いた.

### 3.3 まとめ

強化学習という枠組みを崩さず, 文圧縮という行動を加える事によって文圧縮を考慮に入れる事が出来る

事を見た. この例においては, 最後に抽出された文のみを考慮して圧縮を行なっている. 実際はその文の要約内における依存関係等を考慮して文圧縮が行われる事等も考えられ, そのような既存の要約手法も一つの行動として認識しこの枠組みに落とすことが出来る.

## 4 実験

### 4.1 データセット

DUC2004<sup>1</sup> における Task 2 のデータセットを使って実験を行った. これは, 50 トピックの複数文書要約タスクである. それぞれのトピックの文書群に対して 10 件のニュース文書が含まれる. 要約長の上限は  $K = 100$  と設定する.

### 4.2 実験環境

初期状態  $s_0$  から開始し現在の状態が終了状態  $S_G$  に含まれるまでを 1 エピソードとしたエピソード的タスクとする.

目的スコア関数  $\text{score}(S)$  は, 次のように定義する.

$$\text{score}(S) = \lambda_s \text{Sim}(S, D) - (1 - \lambda_s) \max_{x_i, x_j \in S, i \neq j} \text{Sim}(x_i, x_j) \quad (14)$$

$\text{Sim}$  はコサイン類似度であり, ただし各ベクトルは  $D$  における **tf-idf** 表現ベクトルの値の上位  $d$  要素までのみを考慮するとする. 第一項は要約における妥当性 (relevance), 第二項は冗長性 (redundancy) を表しており,  $\lambda_s$  は  $0 \leq \lambda_s \leq 1$  を満たすトレードオフ定数である.

状態を素性ベクトル  $\phi(s)$  に変換した上で線形手法  $V(s) = \theta^T \phi(s)$  を用いて状態価値関数を表現し, **TD**( $\lambda$ ) により学習する. 素性ベクトル  $\phi(s)$  は, 以下のように要約についてのみ考えた素性を持つ  $d+3$  次元ベクトルとする.

1.  $S$  の **tf-idf** 表現ベクトル
2.  $S$  に含まれる文の, 文位置の逆数
3.  $S$  の文数 / 10.0
4.  $S$  の語数 / 100.0

<sup>1</sup><http://www-nlpir.nist.gov/projects/duc/guidelines/2004.html>

行動選択にはボルツマン選択を用い、初期温度定数  $\tau_0 = 1.0$  とし、1 エピソード毎に 0.989 をかけて減衰させた。学習率は  $\alpha_0 = 0.1/(d+3)$  として  $\alpha(t) = \alpha_0 \cdot 101/(100 + t^{1.1})$  のオーダーで減衰させた。割引率  $\gamma = 1.0$ 、トレース減衰パラメータ  $\lambda = 0.9$  とした。これは強化学習においてよく用いられる形である。

### 4.3 結果

2 章で説明した文抽出のみを行うものを RL、3 章で説明したような行動の追加を行なって文圧縮を考慮したものを RLcomp、DUC2004 におけるベースラインを BASELINE とし、この 3 つの出力に対して ROUGE スコアを用いて評価を行った<sup>2</sup>。ただし、RLcomp においての文圧縮器は実際に実装したものではなく、文をカンマで区切った断片の中でもっとも元文書と類似度が高い一節を返すという実験的なものである。この実験の結果を表 1 にまとめた。RL が BASELINE よりも良く、RLcomp が RL よりも改善されている事が確認され、文圧縮による影響が出ている事が確認できた。

問題点としては、局所解に陥りやすい点がある。終盤になるにつれ収束することは確認できるが、試行回数が少ないと序盤の効果がより大きく影響してしまう。温度定数についての設定と、試行回数について見直す必要がある。また、素性空間が大きくなるほど収束に時間を要する点と、本実験では簡単のため  $V(s) = \theta^T \phi(s)$  の形で実験を行ったが、スコア関数をうまく表現出来るとは限らない点である。これは強化学習においても一つの課題であり、このような問題をより考慮した手法を適用する必要がある結果であった。

## 5 おわりに

自動要約問題を強化学習問題として定式化し、重要文抽出問題のある種の関数近似を加えて解くことが

	ROUGE-1	ROUGE-2	ROUGE-L
RLcomp	0.35250	0.08176	0.30417
RL	0.33937	0.07207	0.28596
BASELINE	0.31524	0.06156	0.27762

表 1: DUC2004 における F スコアの平均。学習エピソードは 300 とし、 $d = 50$ ,  $\lambda_s = 0.9$  とした。提案手法については 10 回の結果の平均値とした。

<sup>2</sup>使用したオプションは、DUC2004 の評価法に則り ./ROUGE-1.5.5.pl -a -c 95 -l 100 -m -n 4 -w 1.2 とした。

できる事を確かめた。次に、既存の文圧縮手法を行動として追加することによって文圧縮についても考慮することが出来る事を確認した。強化学習は試行錯誤によって環境に適応する枠組みであり、今の要約状態でのどのような行動をとればよいかという自然なモデル化が出来たと考える事が出来る。

今後の課題としてはまず、文圧縮パラメータ  $\mu$  は連続値を取る場合も考えられるため、連続値行動を取れるような枠組みの適用を考える必要がある。また、既存の文圧縮手法を取り込み、評価関数の表現の方法を検討し、既存要約手法との性能比較をする事が今後の課題としてあげられる。

## 参考文献

- [1] H.P. Edmundson. New methods in automatic extracting. *Journal of the ACM (JACM)*, Vol. 16, No. 2, pp. 264–285, 1969.
- [2] J. Goldstein, V. Mittal, J. Carbonell, and M. Kantrowitz. Multi-document summarization by sentence extraction. In *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization-Volume 4*, pp. 40–48. Association for Computational Linguistics, 2000.
- [3] H. Jing and K.R. McKeown. The decomposition of human-written summary sentences. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 129–136. ACM, 1999.
- [4] K. Knight and D. Marcu. Statistics-based summarization-step one: Sentence compression. In *Proceedings of the National Conference on Artificial Intelligence*, pp. 703–710, 2000.
- [5] R. McDonald. A study of global inference algorithms in multi-document summarization. *Advances in Information Retrieval*, pp. 557–564, 2007.
- [6] R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*, Vol. 28. Cambridge Univ Press, 1998.
- [7] 高村大也, 奥村学. 最大被覆問題とその変種による文書要約モデル. 人工知能学会論文誌, Vol. 23, No. 6, pp. 505–513, 2008.