

コピー機構を用いたクエリ指向ニューラル生成型要約

石垣 達也^{1,a)} Hen-Hsen Huang^{2,b)} Hsin-Hsi Chen^{2,c)} 高村 大也^{1,3,d)} 奥村 学^{1,e)}

概要: 本研究では、原文書に加えクエリを入力として受け取るクエリ指向要約課題を扱う。ニューラルネットワークが本課題に適用される以前の要約器においては、原文書とクエリでの重複語を要約により含めることで、性能が向上することが報告されている。本研究では、コピー機構を用いて原文書とクエリでの重複語をより要約に含めやすくする仕組みをニューラルネットワークに基づく生成型要約器において実現する。クエリ指向設定に適用可能なコピー機構として、原文書からコピーするモデル、原文書とクエリでの重複語をコピーするモデル、重複語とその周辺語をコピーするモデルを提案する。実験より、重複語とその周辺語をより要約に含めるよう設計したコピー機構が良い性能を示したことを報告する。この結果はニューラルネットワーク以前の要約器において、効果的であった重複語を含める方策が、ニューラルネットワークに基づく要約器においても有効であることを示すものである。

Query-biased Neural Network-based Summarizer using Copying Mechanism

ISHIGAKI TATSUYA^{1,a)} HEN-HSEN HUANG^{2,b)} HSIN-HSI CHEN^{2,c)} TAKAMURA HIROYA^{1,3,d)}
OKUMURA MANABU^{1,e)}

1. はじめに

本研究ではクエリ指向要約課題を扱う。この課題では、表1に示すように、入力として原文書に加えクエリを受け取り、要約を出力する。原文書のみを入力とする一般の単一文書要約課題とは異なり、例えば検索エンジンの検索結果として表示されるスニペットなどに用いることを想定している。古典的な手法としては Tombros ら [18] の抽出型手法などが存在し、クエリと原文書で重複する単語を要約に含めることで、高品質なクエリ指向要約が出力可能なことが知られている。既存手法の多くは抽出型である一方、2017年以降登場している、ニューラルネットワークに基づくクエリ指向要約器 [7], [13] は生成型であり、圧縮率の面で利点がある。しかし、既存のクエリ指向ニューラル要約

原文書:

The “natural death” alternative to euthanasia is not keeping someone alive via life support until they die on life support. That would, indeed, be unnatural. The natural alternative is, instead, to allow them to die off of life support.

クエリ:

Is euthanasia better than withdrawing life support?

要約:

The alternative to **euthanasia** is a natural death without **life support**.

表1 原文書、クエリおよび要約例。要約中の太字で示す単語は原文書とクエリで共に出現している。

器は、クエリと原文書で重複する単語を積極的に含めるといったモデル化をしていない。そこで本研究では、ニューラルネットワークに基づくクエリ指向要約器において、クエリや原文書中の単語を出力に含め易くする枠組みをコピー機構を用いて実現する。

コピー機構はニューラルネットワークに基づく生成型要約器において、原文書中の単語をより要約に含めるための枠組みである。具体的には、出力層での単語の生起確率分布計算時に、原文書に含まれる単語の確率を相対的に高くするようモデル化する。これにより、原文書中の単語が出

¹ 東京工業大学科学技術創成研究院

² 国立台湾大学

³ 産業技術総合研究所

^{a)} ishigaki@lr.pi.titech.ac.jp

^{b)} hhuang@nlg.csie.ntu.edu.tw

^{c)} hhchen@ntu.edu.tw

^{d)} takamura@pi.titech.ac.jp

^{e)} oku@lr.pi.titech.ac.jp

力要約に含まれ易くなる。さらに、コピー機構には事前に定義した単語辞書に含まれない原文書中の単語も出力できる利点もある。このような背景から、既存の単一文書要約に関する研究において、コピー機構は ROUGE 値の向上に大きく寄与することが報告されている [5], [17]。しかしながら、クエリ指向要約課題は原文書に加えクエリを入力として受け取る点において既存の設定と異なる。そのため、コピー機構がクエリ指向要約課題の設定においても有効であるかや、どのようにコピー機構を設計すればより良い性能を得られるかが明らかでない。

本研究では、原文書からコピーする機構 (4.1 節)、クエリと原文書で重複する単語からコピーする機構 (4.2 節)、クエリと原文書で重複する単語とその周辺語からコピーする機構 (4.3 節) を提案する。Nema ら [13] のベンチマークデータセットを用いた実験において、クエリと原文書で重複する単語集合からコピーする機構がより良い ROUGE 値を示したことを報告する。この結果は、ニューラルネットワークを用いる以前の、古典的なクエリ指向要約器において有効性が示されていた、クエリと原文書で重複する単語を要約に含める手法がニューラルネットワークに基づく要約器においても有効であることを示すものである。

2. 関連研究

クエリ指向要約は長らく取り組まれている要約課題の一つである [3], [4], [16], [18]。単一文書もしくは複数文書を入力とする課題設定の双方が存在する。本研究では、単一文書を入力とする設定を扱う。

既存手法のほとんどは抽出型モデルである。これらの手法では、原文書とクエリで重複する単語を重要と考え、このような語を要約に含めることで性能が向上することが示されている [14]。

一方、2017 年以降には、ニューラルネットワークに基づく要約器が抽出型要約および生成型要約の双方において、良い性能を示している [1], [9]。Rush ら [15] はヘッドライン生成課題に対しフィードフォワードニューラルネットワークを用いた要約器を提案し、ニューラルネットワークに基づく手法の有効性を示した。近年提案されている文書要約課題向けの手法の多くは、リカレントニューラルネットワーク (RNN) を用いて構築されている [2], [11], [12]。

ニューラルネットワークに基づく生成型要約器には、コピー機構を備え持つモデルがいくつか提案されている。コピー機構は *Pointer Network* [19] を拡張したモデルと捉えることができる。*Pointer Network* はリカレントニューラルネットワークを用いる要約器において、注意機構が高いスコアを与える原文書中の単語を要約に含める。この仕組みは、原文書に含まれる単語のみを用いて要約を出力する。コピー機構は原文書に含まれる単語だけでなく、事前定義した辞書に含まれる単語からも単語を出力できるよう、

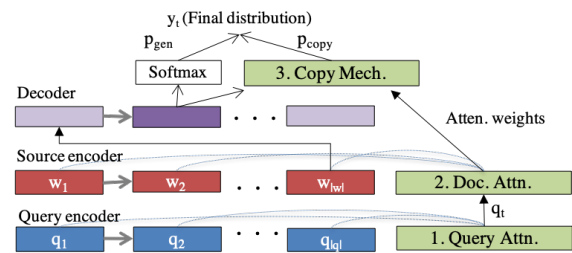


図 1 コピー機構付きクエリ指向ニューラル生成型要約器のモデル概要。

コピーと生成を切り替える機構を持つ [5], [6], [10]。See ら [17], Chen と Lapata [2] は単一文書要約課題において、コピー機構は ROUGE 値を多く向上させることを報告している。これらのコピー機構はクエリを入力としない設定を想定しており、そのままクエリ指向の設定に適用することは出来ない。

クエリ指向設定を想定したニューラルネットワークに基づく要約器も存在する。Hasselqvist ら [7] は原文書から単語をコピーする機構を備えるクエリ指向型要約器を提案した。我々は原文書からクエリで重複する語を優先的にコピーする機構、重複語の周辺語をコピーする機構を新たに提案する。Nema ら [13] は生成型要約課題において出力単語の多様性を高める注意機構を提案した。Debatepedia から構築したクエリ指向型要約のためのデータセットを用いその有効性を示した。本研究でも彼らの構築したデータセットを実験に用いる。

3. 基本モデル

本節では Nema ら [13] がベースラインに用いたコピー機構を持たないニューラルネットワークに基づくクエリ指向要約器について説明する。このモデルを基本モデルとし、4 節に示すコピー機構付きモデル (図 1) に拡張する。基本モデルのエンコーダはクエリを読み込むクエリエンコーダおよび原文書を読み込む原文書エンコーダの 2 つから構成される。デコーダでは、原文書を読み込むエンコーダの生成した隠れ状態に対し、注意機構を用いながら逐次単語の生起確率を計算する。以下各構成要素について順に説明する。

3.1 エンコーダ

クエリ q および原文書 w はそれぞれ単語系列 $q = \{q_1, \dots, q_{|q|}\}$ および $w = \{w_1, \dots, w_{|w|}\}$ と表現する。これらの単語はあらかじめ d -次元の埋め込みベクトルで表現されているものとする。なお、クエリと原文書で単語の埋め込み層は共有する。

クエリを読み込むエンコーダは、以下のように両方向 *Long Short-Term Memory (Bi-LSTM)* に基づくりカレントニューラルネットワークによって、クエリ中の単語を逐次読み

込む: $\vec{h}_i^q = LSTM_q(h_{i-1}^q, q_i)$, $\overleftarrow{h}_i^q = LSTM_q(h_{i+1}^d, q_i)$.
これらの結果を用い、クエリ中の各単語は以下の l 次元ベ
クトル h_i^q で表現される。

$$h_i^q = [\vec{h}_i^q; \overleftarrow{h}_i^q] \quad (1)$$

ここで、 $[\cdot]$ は結合ベクトルを表す。

原文書を読み込むエンコーダも同様に *Bi-LSTM* を用
いて、単語を逐次読み込む: $\vec{h}_j^d = LSTM_s(h_{j-1}^d, q_j)$,
 $\overleftarrow{h}_j^d = LSTM_s(h_{j+1}^d, q_j)$. 原文書中の各単語も同様に *Bi-*
LSTM の出力ベクトルを結合することで、文脈を考慮した
 l 次元ベクトル h_j^d に変換する。

$$h_j^d = [\vec{h}_j^d; \overleftarrow{h}_j^d] \quad (2)$$

原文書を読み込むエンコーダの最終状態のベクトル
和をデコーダ側 *LSTM* の初期状態 s_0 として用いる:
 $s_0 = \vec{h}_{|w|}^d + \overleftarrow{h}_1^d$.

3.2 クエリを考慮する注意機構付きデコーダ

デコーダは *LSTM*, クエリ注意機構, 原文書注意機構お
よび出力層から構成される。

はじめに、クエリ注意機構が時刻 t におけるデコーダ側
LSTM の出力 s_t を用いて、クエリ中の単語 q_i への注意ス
コア $a_{t,i}^q$ を計算し、ソフトマックス関数を用いて確率分布
に変換する:

$$a_{t,i}^q = v_q \cdot \tanh(W_q s_t + U_q h_i^q), \quad (3)$$

$$\alpha_{t,i}^q = \frac{\exp(a_{t,i}^q)}{\sum_{i=1}^{|q|} \exp(a_{t,i}^q)}. \quad (4)$$

ここで、 $W_q, U_q \in \mathbb{R}^{l \times l}$ および $v_q \in \mathbb{R}^l$ である。 $a_{t,i}^q$ を用
いてクエリを表現するベクトル q_t を以下のような荷重合
計ベクトルとして得る:

$$q_t = \sum_{i=1}^{|q|} \alpha_{t,i}^q h_i^q. \quad (5)$$

原文書注意機構は q_t を考慮しながら、さらに入力文書
中のどの単語に着目するかを、以下のように重み $a_{t,j}^d$ として
計算する:

$$a_{t,j}^d = v_d \cdot \tanh(W_d s_t + U_d h_j^d + Z q_t), \quad (6)$$

$$\alpha_{t,j}^d = \frac{\exp(a_{t,j}^d)}{\sum_{j=1}^{|w|} \exp(a_{t,j}^d)}. \quad (7)$$

ここで、 $W_d, U_d, Z \in \mathbb{R}^{l \times l}$ および $v_d \in \mathbb{R}^l$ である。クエリ
を考慮した文書表現 d_t を以下のように荷重合計して得る:

$$d_t = \sum_{j=1}^{|w|} \alpha_{t,j}^d h_j^d. \quad (8)$$

文書表現 d_t および現在の時刻 t のデコーダ *LSTM* の
出力 $s_t = LSTM_d(s_{t-1}, [y_{t-1}; d_{t-1}])$ から、辞書 $N =$

$\{n_1, \dots, n_{|N|}\}$ 中の単語 n に対し重み $score_{t,gen}(n)$ を計算
し、ソフトマックス関数を用いて生起確率分布 $y_{t,gen}$ を
得る:

$$score_{t,gen}(n) = \delta_n \cdot W_o(W_{dec} s_t + V_{dec} d_t), \quad (9)$$

$$p_{t,gen}(n) = \frac{\exp(score_{t,gen}(n))}{\sum_{m=1}^{|N|} \exp(score_{t,gen}(n_m))}. \quad (10)$$

ここで、 $W_o \in \mathbb{R}^{l \times N}$, $W_{dec} \in \mathbb{R}^{l \times l}$, $V_{dec} \in \mathbb{R}^{l \times l}$ である。
式 (9) の $W_o(W_{dec} s_t + V_{dec} d_t)$ により、辞書と同サイズの
ベクトルが得られる。このベクトルの各次元は辞書中の単
語に対応している。 $\delta_n \in \{0, 1\}^{|N|}$ は、単語 n に対応する
次元のみ 1 を取り、それ以外では 0 をとるベクトルである。
辞書サイズのベクトルと δ_n の内積を取ることで単語 n を
要約に含めるスコア $score_{t,gen}$ を算出し、ソフトマックス
関数で確率分布に変換する。得られた確率分布をもとに、
ビーム探索により $p_{t,gen}$ が高くなる単語系列を探索し要約
 y とする。

3.3 目的関数

すべての重み行列は以下のように、目的関数によって、学
習データ D に含まれる正解要約 y に対する負の対数尤度
が最小になるよう更新される:

$$-\frac{1}{|D|} \sum_D \log p(x|y). \quad (11)$$

4. コピー機構の導入

本節では、本研究で提案するクエリを考慮したコピー機
構について述べる。コピー機構は、クエリを入力としない
従来の単一文書要約課題においては、その有効性が示され
ている [5], [17]. 本研究ではクエリを考慮するコピー機構
として、原文書からコピーする機構 (*SOURCE*), クエリと原
文書で重複する単語からコピーする機構 (*OVERLAP*) およ
びクエリと原文書で重複する単語とその周辺語からコピー
する機構 (*OVERLAP-WIND*) を提案する。

4.1 SOURCE: 原文書からコピー

原文書に出現する単語の生起確率を相対的に高くする
ようモデル化するコピー機構はもっとも単純な拡張であ
り、Hasselqvist ら [7] のクエリ指向要約器でも採用されて
いる。原文書からコピーする機構では、事前に定義した
辞書中の単語集合 N と原文書中の単語集合 M の和集合
 $N \cup M = \{n_1, \dots, n_{|N|+|M|}\}$ に対し、時刻 t における単語
生起確率 $p_t(n)$ を考える。

コピー確率 sw_t を考え、式 (10) の $p_{t,gen}$ および以後説明
する $p_{t,copy}$ を重み付け、最終的な単語 n の生起確率 $p_t(n)$
を得る:

$$sw_t = \sigma(z_d \cdot d_t + z_s \cdot s_t + z_y \cdot y_{t-1}), \quad (12)$$

$$p_t(n) = sw_t p_{t,gen}(n) + (1 - sw_t) p_{t,copy}(n). \quad (13)$$

ここで, $z_d, z_s \in \mathbb{R}^l$ および $z_y \in \mathbb{R}^d$ である. σ はシグモイド関数を表す.

このコピー機構において $p_{t,copy}$ を原文書中の単語を要約にコピーする確率とみなし, 以下のように計算する:

$$p_{t,copy}(n) = \begin{cases} \alpha_{t,idx-s(n)}^d & (n \in M), \\ 0 & (n \notin M). \end{cases} \quad (14)$$

ここで, $idx_source(n)$ は単語 n の原文書中での位置を返す関数である. この式では, 原文書に含まれる単語に対しては注意機構の計算したスコア $\alpha_{t,idx-s(n)}^d$ を出力し, 原文書に含まれない単語の場合には値が 0 となる. 式 (6) より, 注意機構はスコアの計算時にクエリに関する情報 q_t も考慮している. よって, $SOURCE$ はクエリを入力としない要約課題向けのコピー機構とは異なり, クエリに対し重要な単語をコピーするようモデル化される.

4.2 OVERLAP: クエリと原文書の重複単語からコピー

クエリと原文書で重複する単語からコピーするモデルを定式化する. この手法では, $SOURCE$ と同様に辞書中の単語集合 N と原文書中の単語集合 M の和集合 $N \cup M = \{n_1, \dots, n_{|N|+|M|}\}$ に対し, 時刻 t における単語生起確率分布 $p_{t,copy}$ を考える. クエリと原文書で重複する単語集合 $N \cap M$ 中の単語についてのみ注意機構のスコアを以下のように増加させる:

$$a_{t,idx-s(n)}^{overlap} = \begin{cases} (1 + \lambda_d) a_{t,idx-s(n)}^d & (n \in N \cap M), \\ a_{t,idx-s(n)}^d & (n \in M \setminus (N \cap M)), \\ 0 & (n \notin M). \end{cases} \quad (16)$$

ここで $idx_s(n)$ は単語 n の原文書中での位置を返す関数である. $a_{t,j}^{overlap}$ をソフトマックス関数により確率分布に変換した値を $p_{t,copy}$ として用いる:

$$p_{t,copy}(n) = \frac{\exp(a_{t,idx-s(n)}^{overlap})}{\sum_{j=1}^{|N|+|M|} \exp(a_{t,j}^{overlap})}. \quad (19)$$

これにより, クエリと原文書で重複する単語に対し相対的に高い生起確率が与えられ, 出力に含まれやすくなる.

4.3 OVERLAP-WIND: クエリと原文書の重複語とその周辺語からコピー

クエリと原文書で重複する単語とその周辺語の生起確率を高くするモデルを定式化する. この手法も $SOURCE$ と同様に $N \cup M$ に対し, 時刻 t における単語生起確率分布 $p_{t,copy}$ を考える. クエリと原文書で重複する単語および重複する語の周辺 $L_d \in \mathbb{N}$ 語のみ注意機構のスコアを以下のように増加させる:

$$a_{t,idx-s(n)}^{wind} = \begin{cases} (1 + \lambda_d) a_{t,idx-s(n)}^d & (n \in M_{L_d}), \\ a_{t,idx-s(n)}^d & (n \in M \setminus M_{L_d}), \\ 0 & (n \notin M). \end{cases} \quad (20)$$

ここで M_{L_d} はクエリと原文書で重複する単語とその周辺 L_d 語からなる単語集合である. ソフトマックス関数により確率分布に変換した値を $p_{t,copy}$ として用いる:

$$p_{t,copy}(n) = \frac{\exp(a_{t,idx-s(n)}^{overlap})}{\sum_{j=1}^{|N|+|M|} \exp(a_{t,j}^{overlap})}. \quad (23)$$

これにより, クエリと原文書で重複する単語およびその周辺語に対し相対的に高い生起確率が与えられ, これらの語が出力に含まれやすくなる.

5. 実験

5.1 データとモデル学習

実験には Nema ら [13] のデータ^{*1}を用いる. このデータは *Debatepedia*^{*2}から抽出したクエリ, 原文書, 要約の組が格納されている. 80%は学習データ, 残りの 20%はそれぞれ 10%ずつ開発データと評価データに分割されている.

ネットワークの重みの最適化には Adam を用いた. 学習率の初期値は 0.0004, $\beta_1 = 0.9$, $\beta_2 = 0.999$ に設定した. 単語埋め込み層は Glove により初期化したのち, モデル本体の学習中にさらにパラメータを更新した. 単語埋め込み層は 300次元に設定した. LSTM の隠れ層次元は 200, 300, 400, バッチサイズは 32, 64 の組み合わせに対し, $L_d = L_q = 0$ および $\lambda_d = \lambda_q = 1.0$ の設定下で開発データの損失関数の値が最小となるよう値を調整した. さらに L_d および L_q は 1, 2 に固定し, λ_d および λ_q を 0.0001, 0.001, 0.01, 0.1, 1.0 の中から開発データでの ROUGE-2 F 値が最大になる設定に調整した. 辞書には訓練データに出現するすべての単語を用い, 入力側および出力側の語彙サイズはそれぞれ 24,577 および 9,555 である.

5.2 比較モデルと評価手法

コピー機構付きのモデル同士の比較に加え 3つのベースライン手法との比較も行う. 1つ目のベースラインは, クエリを考慮しないエンコーダデコーダ (ENC-DEC) で, 式 (6) からクエリを表現するベクトル q_t を除去したモデルである. 2つ目は 3節のクエリを考慮するエンコーダデコーダ (ENC-DEC QUERY) である. 3つ目は Nema ら [13] の提案したクエリ指向要約器 (DIVERSE ATTN) である. このモデルでは, 生成型要約の問題点としてよく知られる単語の繰り返しや情報の多様性が低下する問題に対処するため注意機構を工夫している. これら 3つのベースラインはどれもコピー機構を持たない.

生成型要約課題においては出力長制約が無い. よって, 異なる出力長の要約を適切に評価するため ROUGE-1, ROUGE-2 および ROUGE-L の F 値を用いて評価する.

^{*1} <https://github.com/PrekshaNema25/DiverstiyBasedAttentionMechanism>

^{*2} <http://www.debatepedia.org/en>

	R-1	R-2	R-L
ENC-DEC	13.73	2.06	12.84
ENC-DEC QUERY	29.28	10.24	28.21
DIVERSE ATTN[13]	41.02	26.44	40.78
SOURCE	43.32	29.12	42.96
OVERLAP	43.47	29.68	43.26
OVERLAP-WIND $\{L_d = 1\}$	44.41[†]	30.48[†]	44.20[†]
OVERLAP-WIND $\{L_d = 2\}$	43.16	29.15	42.90
OVERLAP-WIND $\{L_d = 3\}$	44.03	29.78	43.77

表2 ROUGE-F 値.

6. 結果

各比較手法の ROUGE 値^{*3}を表2に示す。表の上部にコピー機構を持たないモデル，下部にコピー機構を持つモデルを示す。

クエリの情報を用いる効果：クエリの情報を考慮しない *ENC-DEC* は極端に低い性能を示した。クエリを考慮する注意機構を導入した *ENC-DEC QUERY* では，ROUGE 値が大幅に向上 (ROUGE-2 において 2.06 → 10.24) した。

コピー機構の効果：原文書からコピーするモデルである *SOURCE* では，コピー機構なしのモデルと比較し大幅に性能が向上 (ROUGE-2 において 10.24 → 29.12) している。また，コピー機構付き既存手法でもっとも良い性能を示している *DIVERSE ATTN* [13] よりも，コピー機構を持つすべてのモデルの ROUGE 値が統計的有意に向上した ($p < 0.05$)。なお，有意差検定には *Paired Bootstrap Resampling* [8] を用いた。よって，本課題においてもコピー機構は有効に機能することが確認できた。

重複語をコピーする効果：OVERLAP は，SOURCE よりも良い ROUGE 値 (ROUGE-2 において 29.12 → 29.68) を示した。OVERLAP-WIND の L_d の値を変化させると $L_d = 1$ の設定においてもっとも高い ROUGE 値を示した。OVERLAP-WIND ($\{L_d\}$) との差は統計的有意であった。これらの結果から，重複する単語をより要約に含めるモデル化はニューラルネットワークに基づく要約モデルにおいても性能を向上させることが分かる。

7. おわりに

本稿ではニューラルネットワークに基づくクエリ指向要約器にコピー機構を導入する手法について述べた。実験より，クエリと原文書で重複する単語集合からコピーする機構が良い性能を示すことがわかった。よって，ニューラルネットワーク以前の手法で有効なクエリと原文書で重複する語を要約に含める手法は，ニューラルネットワークに基づく要約器においても有効であることを報告した。

謝辞：本研究は JST および MOST による国際科学技術協力基盤整備事業「日本-台湾研究交流」のもと行われました。

*3 ROUGE スクリプトのオプションには -a -n 2 -s を用いた。

参考文献

- [1] Bahdanau, D., Cho, K. and Bengio, Y.: *Neural Machine Translation by Jointly Learning to Align and Translate*, Proceedings of ICLR2015 (2015).
- [2] Cheng, J. and Lapata, M.: *Neural Summarization by Extracting Sentences and Words*, Proceedings of ACL2016, Berlin, Germany, pp. 484–494 (2016).
- [3] Dang, H. T.: *Overview of DUC 2005*, Proceedings of DUC2005, Vol. 2005, Citeseer, pp. 1–12 (2005).
- [4] Daumé III, H. and Marcu, D.: *Bayesian Query-focused Summarization*, Proceedings of ACL2006, Association for Computational Linguistics, pp. 305–312 (2006).
- [5] Gu, J., Lu, Z., Li, H. and Li, V. O.: *Incorporating Copying Mechanism in Sequence-to-Sequence Learning*, Proceedings of ACL2016 (2016).
- [6] Gulcehre, C., Ahn, S., Nallapati, R., Zhou, B. and Bengio, Y.: *Pointing the Unknown Words*, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vol. 1, pp. 140–149 (2016).
- [7] Hasselqvist, J., Helmertz, N. and Kågebäck, M.: *Query-Based Abstractive Summarization Using Neural Networks*, arXiv preprint arXiv:1712.06100 (2017).
- [8] Koehn, P.: *Statistical significance tests for machine translation evaluation*, Proceedings of EMNLP2014 (2004).
- [9] Luong, M.-T., Pham, H. and Manning, C. D.: *Effective Approaches to Attention-based Neural Machine Translation*, Proceedings of EMNLP2015, pp. 1412–1421 (2015).
- [10] Miao, Y. and Blunsom, P.: *Language as a Latent Variable: Discrete Generative Models for Sentence Compression*, Proceedings of EMNLP2016, pp. 319–328 (2016).
- [11] Nallapati, R., Zhai, F. and Zhou, B.: *SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents.*, Proceedings of AAAI2017, pp. 3075–3081 (2017).
- [12] Nallapati, R., Zhou, B., dos Santos, C., glar Gulçehre, Ç. and Xiang, B.: *Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond*, Proceedings of CoNLL2016, pp. 280–290 (2016).
- [13] Nema, P., Khapra, M. M., Laha, A. and Ravindran, B.: *Diversity driven attention model for query-based abstractive summarization*, Proceedings of ACL2017, pp. 1063–1072 (2017).
- [14] Otterbacher, J., Erkan, G. and Radev, D. R.: *Biased LexRank: Passage retrieval using random walks with question-based priors*, Information Processing & Management, Vol. 45, No. 1, pp. 42–54 (2009).
- [15] Rush, A. M., Chopra, S. and Weston, J.: *A Neural Attention Model for Sentence Summarization*, Proceedings of EMNLP2015, pp. 379–389 (2015).
- [16] Schilder, F. and Kondadadi, R.: *FastSum: fast and accurate query-based multi-document summarization*, Proceedings of ACL2008(Short paper), pp. 205–208 (2008).
- [17] See, A., Liu, P. J. and Manning, C. D.: *Get To The Point: Summarization with Pointer-Generator Networks*, Proceedings of ACL2017, Vol. 1, pp. 1073–1083 (2017).
- [18] Tombros, A. and Sanderson, M.: *Advantages of query biased summaries in information retrieval*, Proceedings of SIGIR1998, ACM, pp. 2–10 (1998).
- [19] Vinyals, O., Fortunato, M. and Jaitly, N.: *Pointer networks*, Proceedings of NIPS2015, pp. 2692–2700 (2015).