

会話によるニュース記事伝達のための要約

高津 弘明¹ 本田 裕² 藤江 真也^{1,3} 林 良彦¹ 小林 哲則¹

早稲田大学¹ 本田技術研究所² 千葉工業大学³

takatsu@pcl.cs.waseda.ac.jp, Hiroshi_01.Honda@n.t.rd.honda.co.jp,
shinya.fujie@chibakoudai.jp, yshk.hayashi@aoni.waseda.jp, koba@waseda.jp

1 はじめに

BERT (Bidirectional Encoder Representations from Transformers) [Devlin 18] に基づく重要文抽出モデルと文圧縮モデルを提案し、情報量の異なる2種類の要約を主計画として会話させたときの情報伝達効率 (EoIT; Efficiency of Information Transfer) [高津 18] を評価する。

我々はニュース記事のようなまとまった量の情報を効率的に伝達する会話システムの開発を行っている [高津 18]。ここで「効率的」とは、伝達対象となる記事の中から、ユーザーにとって不要な情報を除き、必要な情報だけを伝えることを意味する。我々のシステムの特徴は、あらかじめ主計画、副計画と呼ぶ複数のシナリオを用意しておき、このシナリオに沿って会話を進めることで、リズムの良い会話を実現するうえで必須となる迅速な応答を可能としたところにある。主計画に沿って記事の要点となる情報を提示する傍らで随時ユーザーからのフィードバックを理解し、必要に応じて副計画に遷移して補足情報を提示する。このようにユーザーの興味や理解状態に応じて提示する情報を柔軟に切り替えながら会話を進めていく仕組みを持つ。

一方で、高い情報伝達効率を実現するには、ユーザーの興味や知識レベルに応じた主計画の生成が求められる。例えば、伝達対象のニュースに興味がありそうなユーザーに対しては大きめの主計画を、興味がなさそうなユーザーに対しては小さめの主計画を用意しておくことができれば、より効率的な情報伝達が期待できる。

主計画は、ニュース記事を要約することで作成する。一般的に、要約を作成する手法には、大きく抽出型の手法と非抽出型の手法がある。近年、非抽出型の手法が盛んに研究されているが [Rush 15, Chopra 16, Nallapati 16]、文書の内容と無関係な単語を生成してしまう問題や文書の内容と矛盾する要約を生成してしまう問題がある。このような誤生成問題に取り組んだ研究はあるものの [See 17, Suzuki 17, 清野 18]、実用的なレベルには至っていない。これらの問題に加え、非抽出型の要約モデルの学習には膨大なデータを必要とすること、我々のシステムでは要約で省かれた情報から副計画を作成するため、伝えた情報と伝えていない情報を区別する必要があることを加味し、本研究では抽出型の手法を採用する。

抽出型の手法として、近年では Encoder-Decoder モデルが用いられることが多い [Filippova 15, Cheng 16, Nallapati 17, Al-Sabahi 18]。これらのモデルでは、文や文書全体をエンコードして得られた埋め込み表現を用いて系列ラベリングの要領で文や単語を要約に含めるかどうかを識別する。本研究では、ユーザーの興味に応じて情報量の異なる要約を作成することを目指しているため、重要文抽出においては、抽出する文の数をシステムのハイパーパラメータとして指定できる方が望ましい。そこで、本研究では、単語に重要度を割り当て整数計画法により重要な文を抽出する手法 (最

大被覆モデル) [Gillick 08, Takamura 09] を用いる。最大被覆モデルは、制限要約長に収まる範囲で重要な単語をできるだけ多く被覆するような文集合を抽出する手法である。そのため、単語の重要度の与え方が肝となる。従来、単語の重要度としては、頻度や TF-IDF [Filatova 04]、ロジスティック回帰 [Yih 07, Takamura 09] や RandomForest [高津 18] で推定した値を用いることが多かった。

本研究では、BERT [Devlin 18] を基本構造とした双方向モデルで文脈を考慮して単語の重要度を推定する。これにより、従来の文脈を用いない手法よりも重要文抽出の性能が向上することを示す。また、文圧縮の問題は、文節単位の系列ラベリングの問題として定式化し、単純な双方向 LSTM を用いた場合よりも BERT に基づく提案モデルの方が優れていることを示す。さらに、同論文 [Devlin 18] で提案されている Masked Language Model と Next Sentence Prediction によりモデルの事前学習を行うことで性能が向上することを示す。そして、提案手法で情報量の異なる2種類の要約を作成し、それぞれを主計画として会話させたときの情報伝達効率 [高津 18] について評価した結果を報告する。

2 要約コーパス

日経新聞のテクノロジー系のニュース記事 100 個について 3 人の作業者に要約を作成させた。1 記事あたりの平均文数は 35 文である。要約作業は、重要文抽出、整列、文圧縮の 3 つの作業からなる。

重要文抽出は、ニュース記事の中から重要な文を数文抽出する作業である。ここで、重要文を 2 種類定義した。一つは、見出し的な内容・核となる情報を含む数文を抽出したものである。ここで抽出した文をタイプ a とし、タイプ a の文集合からなる要約を要約 A とする。もう一つは、要約 A の内容を補足するような文を数文抽出し、要約 A に加えたものである。ここで抽出した文をタイプ b とし、タイプ a とタイプ b の文集合からなる要約を要約 B とする。作業には、タイプ a の文を 3 から 6 文、タイプ b の文を 2 から 5 文の範囲で抽出するように指示した。各作業者が選択した各タイプの重要文の平均文数を表 1 に示す。また、各作業者が重要文 a として選んだ文の位置の分布を図 1 に、重要文 b として選んだ文の位置の分布を図 2 に示す。

整列は、抽出した重要文を会話で伝えるのに適した順番に並び替える作業である。作業には、落ちや教訓、今後の展望など会話の締めとなる情報が最後にくるような構成にするように指示した。

文圧縮は、抽出した重要文を会話で伝えるのに適した情報量を削減する作業である。ここでは、各文について会話でこう話したいという口語表現を最初に考えさせ、その口語表現を実現するのに必要な情報を文節単位で選択させた。

表 1: 各作業者が選択した重要文タイプごとの平均文数

	作業 者 1	作業 者 2	作業 者 3
重要文 a	4.15	3.64	3.60
重要文 b	2.61	3.62	3.66

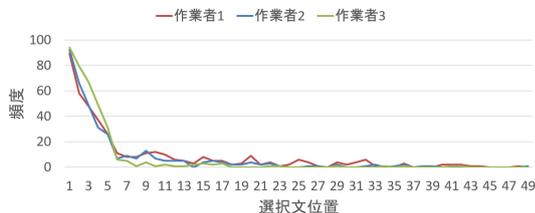


図 1: 各作業者が重要文 a として選んだ文の位置の分布



図 2: 各作業者が重要文 b として選んだ文の位置の分布

3 BERT に基づく要約モデル

BERT [Devlin 18] は, Transformer [Vaswani 17] の Encoder 部分をユニットとする双方向 Transformer モデルである. 文の単語をランダムにマスクし, そのマスクされた単語を予測する Masked Language Model と二つ文が隣接しているかどうかを予測する Next Sentence Prediction の 2 つのタスクで事前学習したモデルを転移学習させることで, 自然言語処理の様々なタスクで SOTA を達成し, 汎用的な言語表現を獲得できるモデルとして注目されている.

本研究では, BERT を抽出型の要約タスクに適用し, 事前学習の効果を検証する.

3.1 BERT の事前学習

日経新聞の 200476 個のニュース記事から段落を超えないように隣接文ペアを重複なく抽出した. この内, 700000 文ペアを訓練セット, 37094 文ペアを開発セットとして, Masked Language Model と Next Sentence Prediction の 2 タスクで BERT の事前学習を行った. 語彙には訓練セットにおいて頻度が 7 以上であった 63272 語を用いた. モデルのパラメータは, Transformer のブロック数を $L = 8$, 隠れ層の次元を $H = 256$, self-attention のヘッド数を $A = 8$ に設定した.

3.2 重要文抽出

伝達対象のニュースに興味があり情報欲求の高いユーザーには大きめの主計画を用意し会話で詳しく説明し, 興味がないユーザーには小さめの主計画で要点のみを伝える. このようなユーザーの興味に応じた主計画の切り替えを可能とすべく, 要点となる文 (重要文 a) とそれを補足する文 (重要文 b) の 2 種類の重要文を抽出することを目指す.

重要文抽出のモデルとして最大被覆モデル [Gillick 08, Takamura 09] を用いる. ここでは, 重要文 a と重要文 b は排他的であるため, 先に重要文 a を抽出し, 文集合から重要文 a を除いたうえで, 重要文 b を抽出する. 整数計画問題の制約として, コーパスのほとんどの要約においてニュース記

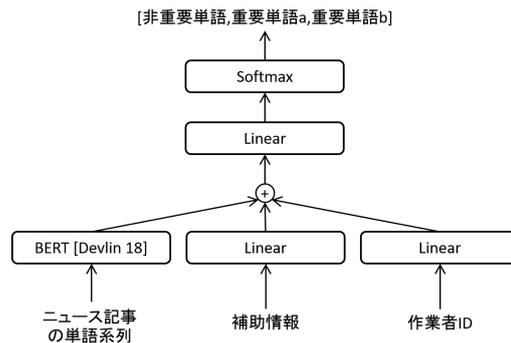


図 3: 単語重要度推定モデル

事の 1 文目が重要文 a として選ばれていたことから (図 1), 重要文 a の抽出においては, 必ず 1 文目を含める制約を加えた.

3.2.1 単語重要度推定モデル

最大被覆モデルは, 制限要約長に収まる範囲で重要な単語をできるだけ多く被覆するような文集合を抽出するモデルである. そのため, 単語の重要度の与え方が肝となる. 従来, 単語の重要度としては, 単語の素性からロジスティック回帰 [Yih 07, Takamura 09] や RandomForest [高津 18] で推定した値を用いることが多かった.

双方向 Transformer の BERT を基本構造として文脈を考慮して単語の重要度を推定するモデルを提案する (図 3). BERT から得られる単語の埋め込み表現に補助情報や作業 ID を加え, 単語が非重要文, 重要文 a, 重要文 b に含まれる確率を softmax で推定する. 補助情報には, JUMAN++¹ (Ver.1.02) の形態素情報 (品詞大分類, 品詞細分類, 活用形, 活用型, カテゴリ, ドメイン), 単語の TF, IDF, TF-IDF, タイトルに含まれるかどうか, 「」内の単語かどうか, 記事における文・段落の位置, KNP² (Ver.4.19) を適用して得られる IREX の 8 種類の固有表現クラス, 係り受けの種類, 係り受け木の深さ, 係り元の文節数, 文頭からの文節位置を用いた. 作業 ID は, 複数の作業者のデータを混ぜて学習させるためのもので, 作業者情報を 1-of-k で区別して与える.

3.2.2 実験設定

2 章で作成した要約コーパスを用いて, 提案モデルによって推定した単語重要度を用いたときの重要文抽出の性能を評価する. 作業 者 i のデータ 90 個を訓練, 作業 者 j のデータ 10 個をテスト ($i \neq j$) となるように分割し, 10 分割交差検定で評価した. 評価尺度には ROUGE-1 を用いた. 抽出する文の数 K は作業者が選択した文の数とした. 整数計画問題を解くための最適化アルゴリズムには分枝限定法を用いた. 被覆する単語の品詞は, 名詞, 動詞, 形容詞, 副詞に限定した.

事前学習なし, 補助情報なしで単一作業者のデータを用いて学習したモデルを SEM1, これに事前学習を行ったモデルを SEM1+pre, さらに補助情報を加えて学習したモデルを SEM1+pre, 作業 者 2 人のデータを用いて学習したモデルを SEM2+pre とする.

比較手法として, 3.2.1 節で説明した補助情報を素性として学習した RandomForest により単語の重要度を与えたときの結果 (RF), 作業 者の一方をシステム出力, 他方を正解として ROUGE-1 を算出した結果 (Human), ランダムに文を抽出したときの結果 (Random), 先頭から K 文抽出したときの結果 (LEAD) と比較を行った.

¹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

²<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

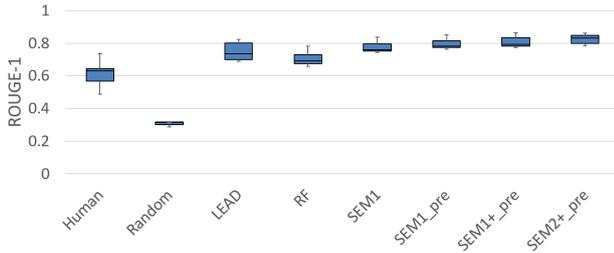


図 4: 重要文 a 抽出の実験結果

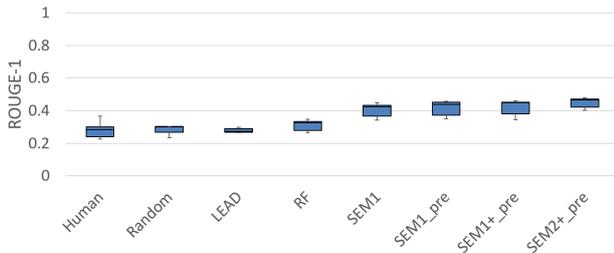


図 5: 重要文 b 抽出の実験結果

3.2.3 実験結果

重要文 a を抽出する実験の結果を図 4 に、重要文 b を抽出する実験の結果を図 5 に示す。重要文 a、重要文 b のいずれの抽出に関しても RandomForest で重要度を与えた場合よりも提案手法の方が良い結果を示した。また、事前学習なしのモデル (SEM1) よりも事前学習を行ったモデル (SEM1_pre) の方が良く、さらに補助情報を加えて学習することで (SEM1+_pre)、複数人のデータを活用することで (SEM2+_pre) 性能が向上することが分かった。

重要文 a の抽出に関しては、LEAD 法が高い性能を示した。これは重要文 a が文の初めの方から選ばれていることに起因する (図 1)。一方で、重要文 b の抽出においては、LEAD 法は Random 以下の性能まで落ち込むことが分かった。

3.3 文圧縮

書き言葉で書かれた文の内容を話し言葉で伝えるのに適した情報量に削減することを目指す。ここで、文圧縮の問題を文節レベルの系列ラベリングの問題として定式化する。BERT に基づく文圧縮モデルを提案し、文法性を保証する枠組みとして野本の手法 [Nomoto 08] を導入する。

3.3.1 文圧縮モデル

双方向 Transformer の BERT に基づく文圧縮モデルを図 6 に示す。BERT で得られた単語の埋め込み表現を self-attention [Lin 17] によって文節ごとに組み上げている点が単語重要度推定モデル (図 3) と異なる。系列ラベリングにおける IOBES ラベルの基準で、文節を省くか含めるかの確率を softmax で推定する。補助情報に関しては、記事における文・段落の位置を除き、3.2.1 で説明したものと同じである。

3.3.2 実験設定

2 章で作成した要約コーパスを用いて、提案モデルの性能を評価する。作業員 i のデータ 90 個を訓練、作業員 j のデータ 10 個をテスト ($i \neq j$) となるように分割し、10 分割交差検定で評価した。

事前学習なし、補助情報なしで単一作業員のデータを用いて学習したモデルを SCM1、これに事前学習を行ったモデルを SCM1_pre、さらに補助情報を加えて学習したモデルを SCM1+_pre、作業員 2 人のデータを用いて学習したモデルを

[文節を省く,含める(開始),含める(内側),含める(終了),含める(単体)]

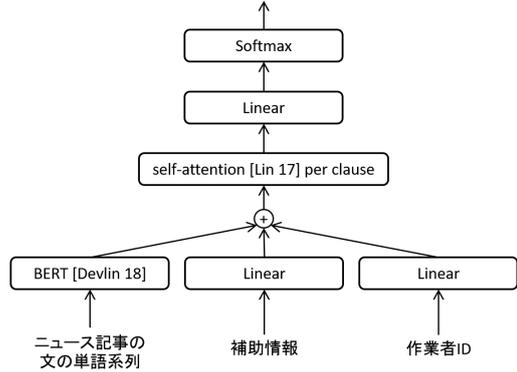


図 6: 文圧縮モデル

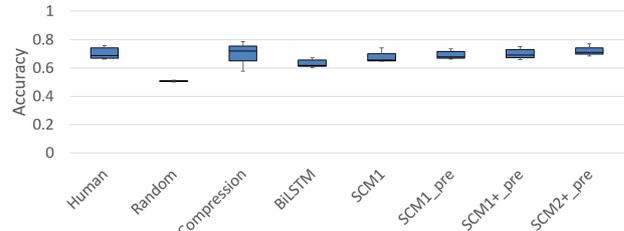


図 7: 文圧縮の実験結果

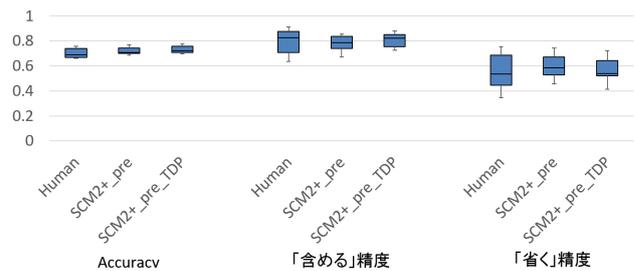


図 8: 文法性を保証した際の文圧縮の実験結果

SCM2+_pre とする。また、係り受け木の刈り込みによってあらかじめ妥当な圧縮文候補を生成しておく [Nomoto 08]、文圧縮モデルで推定した文節の含めるべき確率に基づき各圧縮文候補をランキングし、最上位の結果を採用したときの結果を SCM2+_pre_TDP とする。

比較手法として、3.3.1 節で説明した補助情報と文節を構成する単語の Bag-of-Words を素性として学習した双方向 LSTM により文圧縮を行ったときの結果 (BiLSTM)、作業員の一方をシステム出力、他方を正解として Accuracy を算出した結果 (Human)、ランダムに文節を選択したときの結果 (Random)、文圧縮を行わなかったときの結果 (NoCompression) と比較を行った。

3.3.3 実験結果

Accuracy で評価した結果を図 7 に示す。この結果から提案手法が BiLSTM モデルよりも優れていることが分かった。また、文圧縮のタスクにおいても、事前学習なしのモデル (SCM1) よりも事前学習を行ったモデル (SCM1_pre) の方が良く、さらに補助情報を加えて学習することで (SCM1+_pre)、複数人のデータを活用することで (SCM2+_pre) 性能が向上することが分かった。無圧縮の文の Accuracy (NoCompression) が高いのは、作業員が作成した圧縮文の要約率が 70%程度であることに起因する。

係り受け木の刈り込みによって生成された候補のランキングに基づいて文圧縮を行った結果を図 8 に示す。この結果から、ランキングに基づく手法 (SCM2+_pre_TDP) の方が全

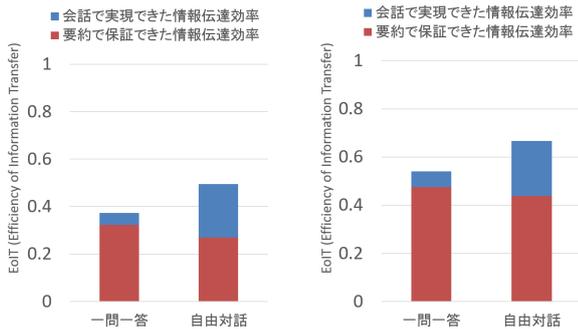


図 9: 情報伝達効率 (要約 A) 図 10: 情報伝達効率 (要約 B)

体の Accuracy は向上するが、文節を「省く」精度において系列ラベリングの結果 (SCM2+_pre) よりもわずかに劣ることが分かった。SCM2+_pre_TDP によって得られる圧縮文の質は係り受け解析の精度に影響されるが、系列ラベリングで含めるべきとされた文節を単純につなげた文よりも文法的に正しいことが多く、要約率の制御も可能であるため実用的であると考えられる。

4 情報伝達効率の評価実験

提案手法で生成した 2 種類の要約 (要約 A と要約 B) を主計画として会話したときの情報伝達効率を評価する。ここで、情報伝達効率とは会話によって提示できたほしい情報の被覆率と知らない情報の除外率の調和平均で定義される尺度である [高津 18]。

4.1 実験設定

日経新聞のテクノロジー系のニュース記事 8 個 (平均 34.6 文) について、4 トピックずつ要約 A と要約 B を作成し、それぞれを主計画としたときの会話実験を 8 人の被験者に対して実施した。ここで、抽出する重要文 a の数を 5、抽出する重要文 b の数を 4 に設定した (要約 A: 5 文, 要約 B: 9 文)。また、本システムを制約なく使用する自由対話タスクと要約を述べた後に QA モードへ移行する一問一答タスク (例えば、吉野らのシステム [Yoshino 15]) について比較を行った。

被験者には、会話終了後にニュース記事を読ませ、各文について会話で伝えて欲しかったかどうかを 5 段階で評価させた (5 点: 伝えてほしい, 4 点: どちらかと言えば伝えてほしい, 3 点: どちらでもよい, 2 点: どちらかと言えばいらない, 1 点: いらない)。会話で提示できた 4 点以上の文の被覆率と会話で提示しなかった 2 点以下の文の除外率の調和平均から情報伝達効率を計算した。

4.2 実験結果

要約 A を主計画として会話したときの情報伝達効率の結果を図 9 に、要約 B を主計画として会話したときの情報伝達効率の結果を図 10 に示す。

この結果から、要約を伝えるだけのシステムや従来の一問一答型の対話システム (一問一答) と比べ、バージョンを許容し即応性高く応答可能な本システム (自由対話) の方が高い情報伝達効率を実現できることが分かった。

また、今回の実験では被験者全体の傾向として知らない情報よりもほしい情報の方が多く見られたため、要約 A よりも要約 B の方が高い情報伝達効率を実現できた。一方で、要約を除き会話のみで実現できた情報伝達効率に着目すると要約 A に基づく会話も要約 B に基づく会話も情報伝達効率に大きな差は見られなかった。

5 おわりに

BERT に基づく重要文抽出モデルと文圧縮モデルを提案し、情報量の異なる 2 種類の要約を主計画として会話させたときの情報伝達効率を評価した。

Masked Language Model と Next Sentence Prediction による事前学習は、抽出型の要約タスクにおいても有効であることが分かった。また、情報欲求の高いユーザーに対して大きめの主計画を設定しておくことで、情報伝達効率が向上することが確認できた。

今回の実験では、被験者に対して要約 A、要約 B をランダムに割り当てて会話させたが、本来であれば、ユーザーのトピックに対する興味度に応じて要約 A と要約 B を使い分けるべきである。今後は、ユーザーとの対話履歴を活用し、ユーザーごとに適切な要約を作成する方法について検討する。

参考文献

[Al-Sabahi 18] Al-Sabahi, K., Zuping, Z., and Nadher, M.: A hierarchical structured self-attentive model for extractive document summarization (HSSAS), *IEEE Access*, Vol.6, pp. 24205-24212 (2018)

[Cheng 16] Cheng, J. and Lapata, M.: Neural summarization by extracting sentences and words, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 484-494 (2016)

[Chopra 16] Chopra, S., Auli, M., and Rush, A.M.: Abstractive sentence summarization with attentive recurrent neural networks, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 93-98 (2016)

[Devlin 18] Devlin, J., Chang, M.W., Lee, K., and Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv:1810.04805*, pp. 1-14 (2018)

[Filatova 04] Filatova, E. and Hatzivassiloglou, V.: A formal model for information selection in multi-sentence text extraction, in *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 397-403 (2004)

[Filippova 15] Filippova, K., Alfonseca, E., Colmenares, C.A., Kaiser, L., and Vinyals, O.: Sentence compression by deletion with LSTMs, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 360-368 (2015)

[Gillick 08] Gillick, D., Favre, B., and Hakkani-Tur, D.: The ICSI summarization system at TAC 2008, *Text Analysis Conference* (2008)

[清野 18] 清野舜, 高瀬翔, 鈴木潤, 岡崎直樹, 乾健太郎, 永田昌明: ニューラルヘッドライン生成における誤生成問題の改善, *言語処理学会第 24 回年次大会発表論文集*, pp. 1-4 (2018)

[Lin 17] Lin, Z., Feng, M., Santos, C.N., Yu, M., Xiang, B., Zhou, B., and Bengio, Y.: A structured self-attentive sentence embedding, in *Proceedings of the 5th International Conference on Learning Representations*, pp. 1-15 (2017)

[Nallapati 16] Nallapati, R., Zhou, B., Santos, C.D., Gulcehre, C., and Xiang, B.: Abstractive text summarization using sequence-to-sequence RNNs and beyond, in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 280-290 (2016)

[Nallapati 17] Nallapati, R., Zhai, F., and Zhou, B.: SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents, in *Proceedings of the 31th AAAI Conference on Artificial Intelligence*, pp. 3075-3081 (2017)

[Nomoto 08] Nomoto, T.: A generic sentence trimmer with CRFs, in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, Human Language Technologies*, pp. 299-307 (2008)

[Rush 15] Rush, A.M., Chopra, S., and Weston, J.: A neural attention model for sentence summarization, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 379-389 (2015)

[See 17] See, A., Liu, P.J., Manning, C.D.: Get to the point: Summarization with pointer-generator networks, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1073-1083 (2017)

[Suzuki 17] Suzuki, J. and Nagata, M.: Cutting-off redundant repeating generations for neural abstractive summarization, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 291-297 (2017)

[Takamura 09] Takamura, H. and Okumura, M.: Text summarization model based on maximum coverage problem and its variant, in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 781-789 (2009)

[高津 18] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図性の異なる多様な情報行動を可能とする音声対話システム, *人工知能学会論文誌*, Vol. 33, No. 1, pp. 1-24 (2018)

[Vaswani 17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., and Kaiser, L.: Attention is all you need, in *Proceedings of the 31st Conference on Neural Information Processing System*, pp. 6000-6010 (2017)

[Yih 07] Yih, W., Goodman, J., Vanderwende, L., and Suzuki, H.: Multi-document summarization by maximizing informative content-words, in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 1776-1782 (2007)

[Yoshino 15] Yoshino, K. and Kawahara, T.: News navigation system based on proactive dialogue strategy, in *Proceedings of the International Workshop series on Spoken Dialogue Systems Technology*, pp. 1-11 (2015)