

会話によるニュース記事伝達のための文圧縮

高津 弘明¹ 福岡 維新¹ 藤江 真也^{1,2} 林 良彦¹ 小林 哲則¹

早稲田大学¹

千葉工業大学²

{takatsu,fukuoka}@pcl.cs.waseda.ac.jp, shinya.fujie@p.chibakoudai.jp,
yshk.hayashi@aoni.waseda.jp, koba@waseda.jp

1 はじめに

ニュース記事のようなまとまりのある情報を会話インタラクションを通じて快適に享受する(受け取って自分のものにする、またその行為を楽しむ)ためには、適度な量の情報を分かりやすい表現でテンポ良く伝えることが重要である [1]。ここでは、特に発話の情報量に着目し、視覚メディアとして書かれたニュース記事の内容を会話メディアに変換して伝える際の発話に含めるべき情報について検討した。そして、会話での利用を目的とした要約コーパスの設計を行い、Differentiable Neural Computers (DNC) をはじめとするいくつかのモデルを用いた文圧縮の実験を行った。

文章による情報伝達では、執筆者が文章を書き終えてから読者がそれを読む。つまり、情報の送信が終わってから、受信が行われることになる。そのため、読者は早く読もうがゆっくり読もうが、一部内容を無視しようが、途中で中断しようが自由である。一方、会話による情報伝達では、話し手が情報を送信すると同時に聞き手が受信することになる(話し言葉の同時性 [3])。そのため、その瞬間ごとに聞き手が受信する情報は常に部分的かつ少量であり、聞き手はその情報の連続体を自分の力でまとめ上げ理解しなくてはならない。しかも、発せられた音声は瞬時に消えてしまうため、文章のように何度も読み返すことができない。このように、会話による情報伝達では、話し手は聞き手に対してコミュニケーション上の大きな負担を強いることになる。

視覚メディアによる情報伝達を前提として書かれたニュース記事の一文一文には、たくさんの情報が埋め込まれている。もしその内容を会話で伝えるのであれば、話し手は送受信の同時性に起因する聞き手の負担を考え、極力情報量を抑え、分かりやすい表現で伝えなければならない。そこで、情報を削減する手段として要約技術、特に、文の内容を短く縮約する手段として文圧縮技術の利用が考えられる。

本稿では、ニュース記事のような視覚メディアによる情報伝達を前提として書かれた文章を会話メディアに変換して伝える際に求められる要約の特殊性について述べた後、会話での利用を目的として設計した要約コーパスについて説明し、最後にそのコーパスを用いた文圧縮の実験結果について報告する。

英国などの 国際研究チームは、 太陽系に 最も 近い 恒星 「プロキシマ・ケンタウリ」の 近くで、 地表に 水が あるかもしれない 地球に 似た 惑星を 発見したと 発表した。
S : 地球に似た惑星が見つかったんだって
U : え?どこで?
S : プロキシマ・ケンタウリっていう恒星の近くだよ
U : 何それ?
S : プロキシマ・ケンタウリっていうのは、太陽系に最も近い恒星なんだよ
U : へー

図 1: 会話利用を目的とした文圧縮の例 1

2 会話利用を目的とした文圧縮

2.1 視覚メディアとしての要約と会話メディアとしての要約の違い

文章を簡潔にまとめる技術として要約技術がある。特に文を対象として短く縮約する技術として文圧縮がある。一般に要約は利用目的に応じて指示的要約と報知的要約に分類できる。指示的要約は、原文を読むかどうかの判断材料としての要約であるのに対し、報知的要約は、原文の代わりとして用いられる要約である。我々が開発している会話システム [1] [2] もユーザーが受け身の姿勢で聞いている限り、ユーザーは要約した内容のみを受け取ることになるが、ここで視覚メディアとしての要約と会話を志向した要約に大きな違いがある。会話で伝えられる要約はたとえ情報が不足していたとしても、送受信の同時性からユーザーはその場で情報要求を示すことで、瞬時に情報を補完できる。一方、報知的要約では、ユーザーが原文を見ることを想定していないため、提示される要約には情報の網羅性が求められる。また、指示的要約では、ある情報をもっと詳しく知りたいと思ったら、原文を参照して該当箇所を探しに行かなければならないのに対し、会話では聞き手からピンポイントで情報を取得することができる。このような送受信の同時性に伴う情報伝達の特徴から、会話では、一発話あたりの情報量は視覚メディアを対象とした要約よりも少なくなる傾向にある。

本研究では、図 1 のような原文が与えられたとき、最初に話すべき内容を文節単位で抽出する。コーパスの

EVスタートアップ Local Motorsが、 IBMの 人工知能 Watsonを 搭載する 自動運転バス 「Olli」を 発表しました。
S : あのさー、Olliって知ってる？
U : えっ、知らない
U : 何それ？
S : Olliっていう自動運転バスがあるんだけど
S : それにね、Watsonが搭載されるんだって
U : えっ、Watsonって、あれ？
U : えっとー、IBMの・・・
S : そうそう、IBMの人工知能

図 2: 会話利用を目的とした文圧縮の例 2

作成にあたっては、作業員に対して各文を会話でどのように伝えたいかを先に考えさせ、次にその口語表現を実現するために必要な情報を文節単位で抽出するよう指示した。

2.2 未知語の説明

文圧縮で抽出が望まれる情報として、核となる情報の他に、知名度の低い用語に対する説明が挙げられる。例えば、図 2 のような原文から「Olli に Watson が搭載されたんだって」という発話が生成されたとすると、ユーザーは「Olli がバス」であるということを知っていなくてはならない。話し言葉の同時性により、ユーザーから「何それ」といった質問がくるかもしれないが、もし、ユーザーが質問せず、知らない用語をそのまま放置してしまつたら、以降の話の理解に支障をきたす恐れがある¹。そのため、知名度の低い用語に対する説明は、なるべく文圧縮の段階で抽出される方が望ましい。逆に「Olli」という単語がニュースを理解する上であまり重要でないとするならば、そもそも「Olli」という単語を出さずに「自動運転バス」と参照してしまつてもよい。その用語に対する説明が不要かどうかはコーパス作成時に作業員が判断する。

3 要約コーパスの設計

ニュースサイトから収集した 100 トピック²について 5 人の作業員が要約を行った。まず、要点となる文を 2~4 文選択させた³。そして、それらの文に対して、まず最初に会話で「こう伝えたい」という表現を作らせた後で、その口語表現を実現するために必要な情報を文節単位で選ばせた (図 1)。なお、文節の区切りは KNP で与えている。

コーパスの統計情報を表 1 に示す。選択された文数は 1 トピックあたり平均 3 文であった。文圧縮における文節数に基づく平均要約率はおよそ 65%程度であつた。

¹我々のシステムでは、後段の処理によっては、システム側から知識確認を行ったり、あえて情報を曖昧にして伝えユーザーからの反応を誘発したりといった対話戦略が発話計画に組み込まれることがある。

²各トピックは 1 つ以上の記事 (5 文以上) から構成され、中には同じ話題について記述した複数記事からなるものも含む。

³「重要文抽出」と「整理」の作業も行っており、これらの作業では、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指示した。

表 1: 選択文数と文圧縮による要約率

	選択文数	要約率 (文節)	要約率 (文字)
作業員 1	297	65.6	74.1
作業員 2	300	65.9	72.0
作業員 3	304	66.5	70.0
作業員 4	302	58.4	65.2
作業員 5	308	69.3	74.4

た。「こう伝えたい」という口語表現と原文との文字数に基づく平均要約率はおよそ 70%程度であつた。後者の方が要約率が高いのは、口語表現では、やさしい表現を作り出すために同格表現や連体修飾節が「~っていう~」「~したんだけど、~」のように展開されることが多く、これにより選ばれた情報量に対して少し長めの表現が作られるためだと考えられる。

4 関連研究

文圧縮は文から情報を削り、文自体を短く縮約するタスクである。統計的な文圧縮の手法として、Knight と Marcu は、Noisy-Channel モデルと決定木による文圧縮手法を提案した [4]。Clarke と Lapata は、文圧縮問題を整数計画問題として定式化し、制限長内で重要度を最大化するような単語群を含めるような文を生成する手法を提案した [5]。野本は、係り受け木の刈り込みにより要約文候補を生成し、CRF を用いて選別する手法を提案した [6]。これらは原文から情報を削る、または、重要な情報を抽出することで文を要約する手法であるが、最近では、ニューラル機械翻訳の手法を応用した生成型の文要約手法が注目を集めている [7] [8]。

本研究は、前者の抽出型のタスクに属し、発話に含めるべき情報を文節単位で抽出する。

5 実験

本研究では、文圧縮の問題を系列ラベリングの問題として捉え、文節単位で発話に含めるか否かを決定する。発話に含める文節にはラベル “I” を付与し、発話に含めない文節にはラベル “O” を付与する。ここでは、モデルとして CRF と双方向 LSTM、Pointer Networks、DNC を用いて比較を行った。

5.1 素性

今回使用した素性を以下に示す。素性ベクトルの次元数はおよそ 1400 次元であつた。

- 文節を構成する文字列
- JUMAN の形態素情報 { 品詞大分類, 品詞細分類, 活用形, カテゴリ, ドメイン }
- IREX の 8 種類の固有表現クラス
- KNP の係り受けの種類 { 係り受け関係, 並列関係, 同格関係, 部分並列内の関係 }
- 係り元の文節数
- 係り受け木における深さ (ルートノードからの距離)
- 文節の位置 / 文の文節総数
- 述語文節が有する格の種類 (e.g. ガ, ヲ, ニ)
- 格の数
- 「」内の文節かどうか

- 『』内の文節かどうか
- 文節を構成する内容語の TF の平均
- 文節を構成する内容語の IDF の最大値
- 文節を構成する内容語の TF-IDF の最大値
- 文節を構成する内容語の重要度の最大値 (訓練セットで要約に含まれるかどうかを学習した RandomForest の確率値)
- 文節内の内容語がタイトルに含まれるかどうか

5.2 Conditional Random Fields

1つ目のモデルとして CRF を用いた。CRF は自然言語処理において形態素解析や固有表現抽出などの系列ラベリングの問題で幅広く使われている。

CRF のツールとして CRF++⁴ を使用した。素性には、上記のものに加え、ウィンドウサイズ 5 以内のバイグラムとトライグラムも加えた。また、過学習を防ぐために L2 正則化を行い、正則化項の係数は 1.0 とした。

5.3 Bidirectional LSTM

2つ目のモデルとして双方向 LSTM (BiLSTM) を用いた。双方向 LSTM は、文節の素性ベクトルを入力として、埋め込み層を経て、順方向と逆方向の隠れ層に伝搬し、同時刻の順方向の隠れ層と逆方向の隠れ層の結合を出力層に伝搬する構造となっている。出力層では softmax を計算して、各ラベルの出力確率を求める。

実験では、埋め込み層と隠れ層を 200 次元に設定した。また、以下 2 つのモデルも含め、実装は Chainer⁵ で行い、最適化アルゴリズムには Adam を用いた。

5.4 Pointer Networks

3つ目のモデルとして Pointer Networks (PN) を用いた。Pointer Networks は、入力の位置に関するポインタを出力とする Encoder-Decoder モデルである。Vinyals らはこのモデルを組合せ最適化問題に適用し、その近似解が得られることを示した [9]。現在研究が進められている生成型要約の手法においても、要約では原文の単語が使われやすいことから、Pointer Networks のような機構をモデルに導入して、出力辞書にない低頻度な単語を入力文章から参照しようという試みがなされている [12]。

オリジナルの Pointer Networks では、Sutskever らのモデル [10] と同様、Encoder から Decoder への入力はシーケンシャルな構造となっている。そこで、Decoder によるより直接的な入力参照を可能とするために、本研究では Bahdanau らのモデル [11] のように Encoder を双方向 LSTM にして attention 機構を導入した。

実験では、Encoder と Decoder の隠れ層ともに 200 次元に設定した。

5.5 Differentiable Neural Computers

4つ目のモデルとして Differentiable Neural Computers (DNC) を用いた。DNC は外部メモリを有するニューラルネットワークである [13]。DNC のメモリには、読み込みヘッドと書き込みヘッドがついており、

ヘッドの位置を動かすことで、ヘッドの位置の情報を読み込んだり、書き込んだりすることができる。これらのヘッドを制御するのが CPU に相当するコントローラである。コントローラには入力ベクトルの他、読み込みヘッドで前時刻に読み込んだ読み込みベクトルも一緒に入力として与えられる。また、出力として出力ベクトルの他、メモリの書き込みに使われる様々なパラメータが出力される。

コントローラとしては様々なタイプのニューラルネットワークを用いて設計することができるが、ここでは LSTM を用いた。実験では、ハイパーパラメータであるメモリスロットの数と次元を 100 とし、読み込みヘッドの数を 5 に設定した。

5.6 実験結果

ある一人の作業者の作業結果を訓練データとしてモデルを学習し、他の作業者の作業結果を評価データとしてモデルの文圧縮結果との F 値を求めた。これを全作業者の組み合わせについて行い、F 値の分布を求め、箱ひげ図で表現したものを図 3 に示す。結果は、DNC が最もよく、続いて Pointer Networks (PN) と双方向 LSTM (BiLSTM) がほぼ同程度の性能を示した。一方、CRF はこれらに比べ、大きく下回る結果であった。これは、LSTM や外部メモリを使用したモデルでは、文節間の長期的な依存関係を捉えられているのに対し、高々数次のマルコフ性しか仮定していない CRF では、こういった長期的な関係性を捉えきれないためだと考えられる。また、ラベル I とラベル O の結果を比較すると、表 1 の要約率から分かる通り、ラベル O の方がラベル I よりもデータ数が少ないため、誤りやすい傾向にある。特に、CRF ではラベルの偏りの影響を受けやすく、O を I と間違えることが多い。そのため、文圧縮の結果が要約率の低いものになってしまう。会話による情報伝達では、いかに情報を削れるかが重要であるため、CRF は目的にそぐわない。実際、CRF の結果を見てみると、長い文ほど誤りやすいことが確認された。一方、DNC では長い文でも削るべき文節を正しく捉えられており、外部メモリの有効性が伺える。

6 おわりに

会話による情報伝達では、文章に書かれた情報と異なり、音声として発せられた情報は瞬時に消えてしまうため、送受信の同時生に伴う聞き手の負担を考慮し、一回の発話で伝える情報は極力抑えなくてはならない。そのため、視覚メディアでの情報伝達を前提として書かれたニュース記事の内容を会話で伝えるためには、情報の削減が必須であり、その手段として文圧縮技術を用いた。実験では、会話での情報伝達を目的とした要約コーパスの設計を行い、DNC をはじめとするいくつかのモデルについて比較し、検討を行った。その結果、外部メモリを有する DNC が最も良い結果を示した。DNC 自体は汎用的なモデルであり、形態素解析や固有表現抽出のような他の系列ラベリングのタスクにおいても有効性が期待できる。一方、Pointer Networks の結果

⁴<https://taku910.github.io/crfpp/>

⁵<http://chainer.org/>

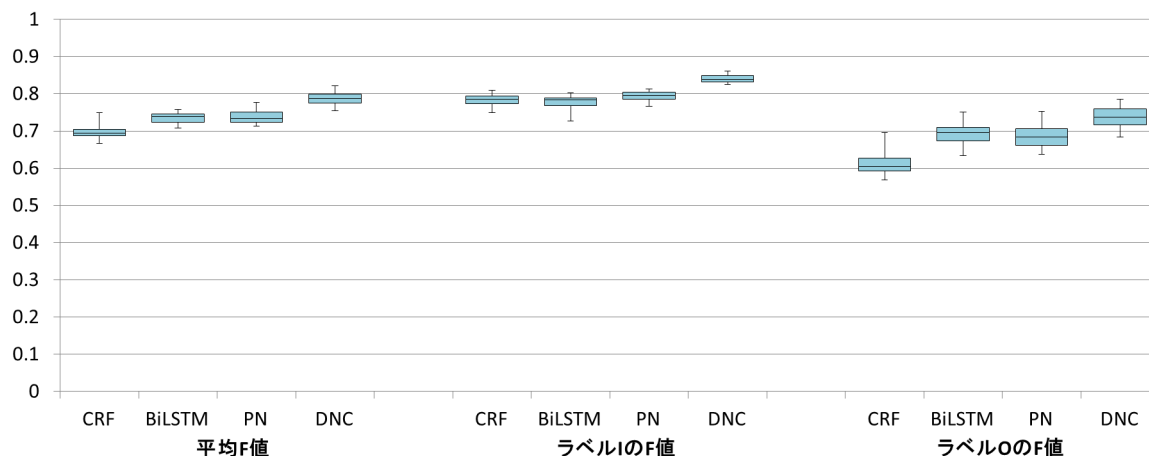


図 3: 実験結果

は DNC には劣るものの、Pointer Networks では、発話に含めるか否かだけでなく、発話における文節の順序関係も学習することができるため、後段の口語化処理での応用可能性についても今後検討していきたい。

本稿では、発話に含めるべき情報の抽出に関してのみ触れたが、聞き手にとって負担の少ない情報伝達（自然な会話インタラクション）を実現するには、情報量の他、その内容をどのような言語表現で伝えるべきかといった口語化 [16] の課題や、どのような速度やイントネーションで伝えるべきかといった音声合成 [17] の課題についても取り組む必要がある。

参考文献

- [1] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: “快適な情報享受を可能とする音声対話システム”, 言語処理学会第 22 回年次大会発表論文集, pp.302-305, 2016.
- [2] 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”, 人工知能学会全国大会, 204-OS-23a-3in2, pp.1-4, 2016.
- [3] 畠弘巳: “話しことばの特徴 - 冗長性をめぐって -”, 国文学解釈と鑑賞, Vol.52, No.7, pp.22-34, 1987.
- [4] K.Knight and D.Marcu: “Summarization beyond sentence extraction: A probabilistic approach to sentence compression”, *Journal of Artificial Intelligence Research*, Vol.139, pp.91-107, 2002.
- [5] J.Clarke and M.Lapata: “Global Inference for Sentence Compression: An Integer Linear Programming Approach”, *Journal of Artificial Intelligence Research*, Vol.31, pp.399-429, 2008.
- [6] T.Nomoto: “A Generic Sentence Trimmer with CRFs”, In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, Human Language Technologies, Columbus, pp.299-307, 2008.
- [7] A.M.Rush, S.Chopra, and J.Weston: “A Neural Attention Model for Sentence Summarization”, In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.379-389, 2015.
- [8] S.Chopra, M.Auli, and A.M.Rush: “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.
- [9] O.Vinyals, M.Fortunato, and N.Jaitly: “Pointer Networks”, In Advances in Neural Information Processing Systems 28, pp.2692-2700, 2015.
- [10] I.Sutskever, O.Vinyals, and Q.V.Le: “Sequence to Sequence Learning with Neural Networks”, In Advances in Neural Information Processing Systems, pp.3104-3112, 2014.
- [11] D.Bahdanau, K.Cho, and Y.Bengio: “Neural Machine Translation by Jointly Learning to Align and Translate”, In Proceedings of the International Conference on Learning Representations, 2015.
- [12] C.Gulcehre, S.Ahn, R.Nallapati, B.Zhou, Y.Bengio: “Pointing the Unknown Words”, In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp.140-149, 2016.
- [13] A.Graves, G.Wayne, M.Reynolds, T.Harley, I.Danihelka, A.G.Barwinka, S.G.Colmenarejo, E.Grefenstette, T.Ramalho, J.Agapiou, A.P.Badia, K.M.Hermann, Y.Zwols, G.Ostrovski, A.Cain, H.King, C.Summerfield, P.Blunsom, K.Kavukcuoglu, and D.Hassabis: “Hybrid computing using a neural network with dynamic external memory”, *Nature* 538, pp.471-476, 2016.
- [14] A.Graves, G.Wayne, I.Danihelka: “Neural Turing Machines”, arXiv preprint arXiv:1410.5401, 2014.
- [15] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: “会話によるニュース記事伝達のための情報選択”, 言語処理学会第 22 回年次大会発表論文集, pp.199-202, 2016.
- [16] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: “会話によるニュース記事伝達のための口語化における述語の書き換え”, 言語処理学会第 22 回年次大会発表論文集, pp.1033-1036, 2016.
- [17] 福岡維新, 高津弘明, 藤江真也, 林良彦, 小林哲則: “会話による情報伝達における発話系列の韻律分析”, 人工知能学会全国大会, 2P1-5in1, pp.1-4, 2016.
- [18] 林良彦, 藤江真也, 福岡維新, 高津弘明, 小林哲則: “情報アクセスにおける受動性と能動性: 音声対話によるニュース記事アクセス”, 人工知能学会第 11 回インタラクティブ情報アクセスと可視化マイニング研究会, pp.25-32, 2015.
- [19] S.Fujie, I.Fukuoka, A.Mugita, H.Takatsu, Y.Hayashi and T.Kobayashi: “A Spoken Dialog System for Coordinating Information Consumption and Exploration”, ACM SIGIR Conference on Human Information Interaction and Retrieval, pp.253-256, 2016.