

複数文質問を対象とした抽出型および生成型要約

石垣 達也[†]・高村 大也^{††,†††}・奥村 学^{†††}

インターネット上のコミュニティ QA サイトや学会での質疑応答の場面などにおいて、人々は多くの質問を投げかける。このような場面で用いられる質問には、核となる質問に加え補足的な情報をも付与され、要旨の把握が難しくなることもある。補足的な情報は正確な回答を得るには必要であるが、質問の要旨を素早く把握したいといった状況においては必ずしも必要でない。そこで本稿では、新たなテキスト要約課題として、複数文から構成される質問テキストを単一質問文に要約する“質問要約”を提案する。本研究ではまず、コミュニティ質問応答サイトに投稿される質問から質問テキスト-要約対を獲得し、必要な要約手法について抽出型および生成型の観点から分析を行う。また、獲得した質問テキスト-要約対を学習データとして抽出型および生成型の要約モデルを構築し、性能を比較する。分析より、抽出型要約手法では要約できない質問テキストの存在を確認した。また要約モデルの比較実験から、従来の要約課題で強いベースライン手法として知られるリード文よりも、先頭の疑問文を規則を用いて同定し抽出するリード疑問文ベースラインがより良い性能を示すこと、生成型手法であるエンコーダ・デコーダモデルに基づく要約手法が、ROUGE による自動評価、人間による評価において良い性能を示すことなどの知見を得た。また、入力中の出現単語を出力に含めるコピー機構を持つエンコーダ・デコーダモデルは、さらに良い性能を示した。

キーワード：文書要約，抽出型要約，生成型要約，ニューラルネットワーク

Extractive and Abstractive Summarization for Multiple-sentence Questions

TATSUYA ISHIGAKI[†], HIROYA TAKAMURA^{††,†††} and MANABU OKUMURA^{†††}

Questions are asked in many situations such as sessions at conferences and inquiries through emails. In such situations, questions can be often lengthy and hard to understand, because they often contain peripheral information in addition to the main focus of the question. Thus, we propose the task of question summarization. In this research, we firstly analyzed question-summary pairs extracted from a Community Question Answering (CQA) site, and found that there exists the questions that

[†] 東京工業大学大学院総合理工学研究科知能システム科学専攻, Department of Computational Intelligence and System Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

^{††} 産業技術総合研究所人工知能研究センター, Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST)

^{†††} 東京工業大学科学技術創成研究院未来産業技術研究所, Laboratory for Future Interdisciplinary Research of Science and Technology (FIRST), Institute of Innovative Research (IIR), Tokyo Institute of Technology

cannot be summarized by extractive approaches, but abstractive approaches are required. We created a dataset by regarding the question-title pairs posted on a CQA site as question-summary pairs. By using the data, we trained extractive and abstractive summarization models, and compared them based on the ROUGE score and manual evaluation. Our experimental results show an abstractive method, the encoder-decoder with the copying mechanism, achieves better scores both on ROUGE-2 F-measure and the evaluation by human judges.

Key Words: *Summarization, Extractive Summarization, Abstractive Summarization, Neural Networks*

1 はじめに

学会での質疑応答や電子メールによる問い合わせなどの場面において、質問は広く用いられている。このような質問には、核となる質問文以外にも補足的な情報も含まれる。補足的な情報は質問の詳細な理解を助けるためには有益であるが、要旨を素早く把握したい状況においては必ずしも必要でない。そこで、本研究では要旨の把握が難しい複数文質問を入力とし、その内容を端的に表現する単一質問文を出力する“質問要約”課題を新たに提案する。

コミュニティ質問応答サイトである Yahoo! Answers¹から抜粋した質問の例を表1に示す。この質問のフォーカスは“頭髪の染料は塩素によって落ちるか否か”である。しかし、質問者が水泳をする頻度や現在の頭髪の色などが補足的な情報として付与される。このような補足的な情報は正確な回答を得るためには必要であるが、質問内容をおおまかに素早く把握したいといった状況においては、必ずしも必要でない。このような質問を表1に例示するような単一質問文に要約することにより、質問の受け手の理解を助けることが出来る。本研究では、質問要約課題の一事例としてコミュニティ QA サイトに投稿される質問を対象テキストとし、質問への回答候補者を要約の対象読者と想定する。

表 1 複数文質問とその要約

質問：
I'm a swimmer for my school swim team and I practice two hours a day, five days a week. I would like to dye my hair black (it is dark brown now) but I am wondering whether the chlorine will stripe it. Will it or will it not ?
要約：
1. Will the chlorine stripe my hair ?
2. Affect of the chlorine on my hair dyeing.

¹ <https://answers.yahoo.com/>

テキスト要約課題自体は自然言語処理分野で長く研究されている課題の一つである。既存研究は要約手法の観点からは、大きく抽出型手法と生成型手法に分けることができる。抽出型手法は入力文書に含まれる文や単語のうち、要約に含める部分を同定することで要約を出力する。生成型手法は入力文書には含まれない表現も用いて要約を生成する。一方で、要約対象とするテキストも多様化している。既存研究の対象とするテキストは、従来の新聞記事や科学論文から、最近では電子メールスレッドや会話ログなどに広がり、それらの特徴を考慮した要約モデルが提案されている。(Duboue 2012; Oya and Carenini 2014; Oya, Mehdad, Carenini, and Ng 2014)

質問を対象とする要約研究としては Tamura, Takamura, and Okumura (2005) の質問応答システムの性能向上を指向した研究が存在する。この研究では質問応答システムの構成要素である質問タイプ同定器へ入力する質問文を入力文書から抽出する。本研究では、彼らの研究とは異なり、ユーザに直接提示するために必要な情報を含んだ要約の出力を目指す。ユーザに直接提示するための質問要約課題については、既存研究では取り組まれておらず、既存要約モデルを質問テキストに適用した場合の性能や、質問が抽出型手法で要約可能であるか、生成型の手法が必要であるか明らかでない。そこで、本研究ではコミュニティ質問応答サイトに投稿される質問テキストとそのタイトルの対（以後、質問テキスト－タイトル対と呼ぶ）を、規則を用いてフィルタリングし、質問テキストとその要約の対（以後、質問テキスト－要約対と呼ぶ）を獲得する。獲得した質問テキスト－要約対を分析し、抽出型および生成型の観点から質問がどのような手法を用いて要約可能であるか明らかにする。また、質問要約課題のために、ルールに基づく手法、抽出型要約手法、生成型要約手法をいくつか構築し性能を比較する。ROUGE (Lin 2004) を用いた自動評価実験および人手評価において、生成型手法であるコピー機構付きエンコーダ・デコーダモデルがより良い性能を示した。

2 質問要約課題

本稿で提案する“質問要約”についてその特徴や既存要約課題との相違について議論する。本研究で扱う質問要約はテキスト要約課題の一つである。既存のテキスト要約研究では、Document Understanding Conference (DUC)²などの新聞記事や科学論文をもとにした共通のデータセットがしばしば性能評価実験に用られてきた。本研究で扱う質問要約は質問テキストを要約対象とし、出力を単一質問文に限定する。

はじめに入力の質問テキストの特徴について述べる。質問要約は従来の新聞記事や科学論文を基に作られたテキストを想定する要約課題とは以下の点において異なる特徴を持つ。

² <https://duc.nist.gov>

- (1) 質問文と叙述文が混在する（個別の文の性質の違い）
- (2) 質問文を叙述文が補足する文間関係を持つ（談話構造の違い）

前者については、新聞記事や科学論文には質問文がほとんど含まれないのに対し、質問テキストでは多くの場合1文以上の質問文を含む。そのため、文書に含まれる文そのものの性質が従来の要約対象とは性質を持つ。後者は文間の関係に着目した相違点である。新聞記事ではリード文を他の文から補足する談話構造がしばしばみられるのに対し、質問テキストでは核となる質問文が存在し、質問文以外の文が核となる質問文を補足する。よって、文そのものの性質の違いに加え、質問テキストは文間の意味的な関係（談話構造）も異なる。このように、個別の文の性質や談話構造に関し従来の要約課題とは異なる特徴を持つことから、例えば従来の要約課題において強いベースライン手法として知られるリード法やその他の既存手法が、質問テキストに対しどの程度の性能を示すかはそもそも明らかではない。

次に出力の単一質問文が満たすべき性質について議論する。本研究では質問テキストの一事例としてインターネット上でのコミュニティ質問応答サイトに投稿される複数文質問の要約を想定する。生成した要約を回答者候補に提示することで、回答者候補は質問内容が回答可能であるか素早く判断できるようになる。このような目的を鑑み、本研究では入力として複数文質問、出力は質問内容を端的に表現した単一質問文を考える。例えば、表1に示す複数文質問に対する正解要約としては、“Will the chlorine stripe my hair?（塩素によって髪の色が落ちますか？）”といった疑問文だけでなく、“Affect of the chlorine on my hair dyeing.（プール中の塩素の染髪への影響）”といった表現も正しい要約と考える。前者は「末尾が?であるか」「先頭の語が助動詞であるか」などの単純な規則を用いて同定できる。後者は質問内容を推測できるが、パターンが無数にあり単純な規則ではこれらを同定できない。本研究ではどちらの表現も要約とみなす。後者まで含んだ広い表現を質問文と呼び、単純な規則で同定可能な前者を疑問文と呼び区別する。

3 関連研究

テキスト要約の既存研究は多く存在する。DUCに代表される多くのタスクなど、既存研究の多くは新聞記事や科学論文を対象としている。近年では、会話テキストや電子メールスレッドを対象とした新たな要約課題も提案されている (Duboue 2012; Oya and Carenini 2014; Oya et al. 2014)。

本研究に関連する取り組みとして、質問応答システムのための質問要約研究が存在する。Tamura et al. (2005) は、複数文で構成される質問を入力として受け付ける、質問応答システムの構築を目指した。この研究では、複数文質問からもっとも核となる1文を抽出する要約器を質問応答システムの前処理として組み込むことで、複数文質問を受け付けるシステムを実現している。

彼らは複数文質問を単一質問文に要約することで、質問応答システムの質問タイプ同定器の性能が向上することを報告している。一方、抽出した核となる質問文は常にユーザの理解できる情報を含むとは限らない。例えば、表1における核文は“Will it or will it not?”であるが、この文には *chroline* や *hair* などといった質問内容を把握するために必要となる単語が含まれず、要約として提示するには情報が不足する。本研究は、ユーザに提示するための要約を出力を指向するため目的が異なる。

抽出型要約モデルの既存研究としては、単語出現頻度を用いて文にスコアを与える手法 (Luhn 1958) や、文同士の類似度を用いて重要文を同定するヒューリスティックを用いる手法 (Mihalcea and Tarau 2004) などが提案されている。さらに、文に対し抽出した場合の ROUGE 値を回帰モデルを用いて予測するモデル (Peyrard and Eckale-Kohler 2016; Li, Qian, and Liu 2013) や、要約に含めるべき文を二値分類する分類問題として定式化する教師あり学習を用いる手法 (Hirao, Isozaki, Maeda, and Matsumoto 2002; Shen, Sun, Li, Yang, and Chen 2007) も存在する。

生成型要約としては、入力の話話構造木を枝刈りする手法 (Dorr, Zajic, and Schwartz 2003; Zajic, Dorr, and Schwartz 2004) や、機械翻訳モデルを用いる手法 (Bank, Mittal, and Witbrock 2010; Woodsend, Feng, and Lapata 2010; Cohn and Lapata 2013), テンプレートを用いる手法 (Oya et al. 2014) などが存在する。近年では、機械翻訳課題向けに提案されたエンコーダ・デコーダモデル (Luong, Pham, and Manning 2015; Bahdanau, Cho, and Bengio 2014) を要約課題に適用する手法 (Rush, Chopra, and Weston 2015; Kikuchi, Neubig, Sasano, Takamura, and Okumura 2016; Gu, Lu, Li, and Li 2016) が積極的に研究されている。

4 質問応答サイトからの質問テキスト-要約対獲得と分析

本研究ではまず、Yahoo! Answers Comprehensive Question and Answers version 1.0³を対象に事例分析を行う。Yahoo! Answers において、ユーザは自由に質問テキストとそのタイトルを記述し投稿する。このデータセットには 4,484,032 の質問投稿が含まれる。質問投稿の中には質問テキスト-タイトル対を質問テキスト-要約対とみなせる事例もあれば、みなせない事例も存在する。そのため、データセット内の質問テキスト-タイトル対を、規則を用いてフィルタリングし、質問テキスト-要約対を獲得する必要がある。その上で、獲得した質問テキスト-要約対に必要な要約手法の検討のための分析および要約モデルの比較実験における学習データに用いる。よって、本研究では以下の二段階でデータセットの分析を行う。

- (1) 質問テキスト-要約対とみなすことのできる質問テキスト-タイトル対の特徴はなにかを明らかにし、質問テキスト-要約対を獲得する。

³ <https://webscope.sandbox.yahoo.com/>

- (2) 質問テキスト-要約対を分析し、抽出型の手法で要約可能か、生成型の手法が必要であるか明らかにする。

4.1 分析 1: 質問の長さ

質問テキスト-要約対とみなすことのできない事例をフィルタリングする規則を設計するために、まず質問に含まれる文数に着目した分析を行った。質問テキストが1文から5文で構成される事例をデータセットからランダムに20事例ずつ抽出し、それらの質問テキスト-タイトル対を質問テキスト-要約対とみなすことができるか否かを人手で判定した。表2に結果を示す。

質問テキストの長さが1文もしくは2文の場合には、質問テキスト-要約対とみなすことのできない事例が増える。このような事例においては、タイトルに核となる質問文が記述され、質問テキスト中では補足的な内容だけが記述され、質問文を含まない例が見られる。質問テキストでは質問文が記述されないため、質問テキスト-タイトル対を質問テキスト-要約対とみなすことができない。一方、3文以上から構成される質問テキストにおいては質問テキスト-要約対とみなすことのできる事例が一定になる。

4.2 分析 2: 質問テキストとタイトルでの名詞の重複

次に質問テキストとタイトルでの名詞の重複に着目した分析を行った。以下に質問テキスト-要約対とはみなすことのできない事例を示す。

タイトル：

Why is there often a mirror in an elevator?

質問テキスト：

I just realized this when I was in an elevator. Does anybody know the reason? What is the history behind it?

この事例では、質問テキスト中の“it”は“エレベータ内に鏡が存在する”という事実を指し示す。“it”が何を指し示すかを理解するには“elevator”や“mirror”といった重要な単語が質問テキスト中に含まれている必要がある。しかし、“mirror”という単語はタイトルには出現する

表 2 質問テキスト中の文数と質問テキスト-要約対とみなせる事例の割合

質問テキスト中の文数	要約対とみなせる事例の割合
1	3/20
2	8/20
3	14/20
4	15/20
5	15/20

が質問テキストには出現しない．そのため，質問テキストを要約器の入力として，タイトルの“Why is there often a mirror...”という要約を生成することはできない．データセット中では，この例のように質問テキストからタイトル中の単語を照応したり，タイトル中の単語が質問テキストを理解するために不可欠である事例を多く観測した．このような事例をフィルタリングするために，我々はタイトルと質問テキストでの単語の重複がフィルタリングのための重要な手がかりとなると考えた．

4.3 分析 3: 抽出型 vs. 生成型

次に必要な要約手法に着目した分析を行った．具体的には質問テキスト－要約対とみなすことのできる事例について，抽出型手法を用いて要約可能であるか，生成型手法が必要であるか人手で分類した．質問テキストの文数が3-5文である20事例を無作為に抽出し，質問テキスト－要約対とみなせるか否かについて分析を行った．

分析の結果を表3に示す．また，代表的な質問テキスト－タイトル対の例を表4に示す．20事例中5事例は質問テキスト－要約対とはみなせない事例であった．20事例中8事例は抽出型手法でタイトルと同等の要約を生成できることがわかった．表4の例2-1および例2-2に抽出型手法により要約可能な事例を例示する．例2-1において，2文目の疑問文を抽出することでタイトルと同等の要約を出力できる．例2-2は先頭文を抽出することでタイトルと同等の要約を生成できる．しかし，実際のタイトルでは“Can someone tell me”といった表現が除去されたり，本文中での“cellular phone promotion”という具体的な表現が“cellular phone plan”というより抽象的な表現に言い換えられている．このように，抽出型手法が適用できる質問テキストであっても実際のタイトルは生成的に作られている場合がある．

残りの7事例については，抽出型手法では適切な要約を出力できず，生成型手法を必要とする．生成型手法が必要な理由としては，例3のように照応解析が必要であったり，例4のように複数文にまたがる情報を適切に埋め込む必要のある事例が存在する．このような事例に対しては，抽出的な手法をただ適用するだけでは，要旨の把握に必要な情報を適切に要約に含めることが出来ない．

表 3 手法検討のための事例分析

質問テキスト－要約対とみなせない	5/20
抽出型手法で要約可能	8/20
生成型手法が必要	7/20

表 4 Yahoo! Answers Comprehensive Question and Answers version 1.0 に含まれる代表的な質問テキスト-タイトル対

例 1 (タイトルが質問テキストの内容を含まない) : I accidentally spilled buttered popcorn on my leather hospital shoe. It has dark spots on it now and I don't know how i can get them off. ... タイトル: Please help!
例 2-1 (抽出型手法によって要約可能で, タイトルも抽出的に作成) : I have not been smoking for six months now, with a small interruption of one week where I smoked like three cigarettes a day, but I did not feel anything anymore then! What is the best way to start smoking again? I am living in an apartment complex that prohibits smoking even inside. タイトル: What is the best way to start smoking again?
例 2-2 (抽出型手法によって要約可能で, タイトルは生成的に作成) : Can someone tell me the best cellular phone promotion in the bay area? A promotion that you run into recently (e.g., local stores)? Or, a web-site promotion that you find (e.g., amazon)? I had a family plan with Cingular and would like to get a good deal when I renew or switch to a new provider. タイトル: Best cellular phone plan in the bay area?
例 3 (生成型手法が必要 1) : "The Simpsons" is one of the funniest shows ever . It's one of my favorites . Do you like it ? タイトル: Do you like "The Simpsons" ?
例 4 (生成型手法が必要 2) : I want my chocolate chip cookies to be thicker and kind of gooey-crispy outside, chewy inside. I've experimented with various recipes and various oven temperature, but my cookies always turn out thin and flat. Why? What am I doing wrong? タイトル: Why do my chocolate chip cookies always turn out thin and flat?

5 データと比較手法

本節では以上の分析をもとに, データセットに含まれる質問テキスト-タイトル対をフィルタリングし, 抽出型および生成型の要約モデルを実際に構築し質問要約課題に適用した場合の性能を比較する.

5.1 データセット

5.1.1 規則によるフィルタリング

分析に基づき設計した以下の条件を満たす事例をフィルタリングし, 質問テキスト-要約対とみなすことのできない事例を除外する. 表5に示すように, Yahoo! Answers データセットにはタイトルのみが記述され質問テキストが記述されない事例が多く存在する. データセット全

表 5 Yahoo! Answers dataset 全体に含まれる質問テキストの長さ

質問テキスト中の文数	0	1	2	3	4	5	6	7 以上
文書数	2,291,554	774,108	434,563	297,158	202,240	136,163	92,760	354,562

0 文はタイトルのみが記述され、質問テキストが存在しない投稿を示す。

体の 4,483,031 投稿のうち、質問テキストおよびタイトルの双方が記述された 2,191,477 投稿について以下の規則を逐次適用し、フィルタリングを行い、質問テキスト-要約対を獲得する。

タイトルが複数文で構成される タイトルが 2 文以上で構成される事例

短い質問テキスト 質問が 2 文以下である事例

単語の重複 タイトル中の名詞が本文中に出現しない事例

短いタイトル, 長いタイトル タイトルが 3 単語以下、16 単語以上の事例

これらの規則を設定した根拠について補足する。複数文から構成されるタイトルを含む投稿をフィルタリングするのは、本課題が単一質問文への要約課題であることがその理由である。短い質問テキストによるフィルタリングは、質問テキスト-要約対となりやすいのは 3 文以上の質問テキストであることであるという 4 節での分析に基づく。単語の重複についてのフィルタリングも同様に、4 節での分析に基づいている。これらの分析に基づくフィルタリングに加え、「短いタイトル」および「長いタイトル」によるフィルタリングも用いる。もっとも単純な疑問文の構文である “What is xx ?” や “Is this xx ?” は最低でも 4 単語必要とする。そのため、3 単語以下のタイトルには必要な情報が含まれていないと考えた。タイトルが 3 語以下で構成される質問テキスト-タイトル対をランダムに 20 事例抽出し調査すると、抽出したすべてのタイトルで “difficult question?”, “Question on Pokemon” といったように、回答者が回答可能であるか判断するための情報が欠如していた。よって、短いタイトルを含む事例はフィルタリングすることにした。Yahoo! Answers において記述される 15 単語以下で記述されるタイトルは全体の 81% を占める。残りの 19% の部分に関してはタイトルが長く圧縮率が低いと考えフィルタリングすることにした。

なお、タイトルが疑問文であるかといった手がかり語などの規則を用いたフィルタリングは行わない。タイトルの表現には “How the Bordeaux Classification was born ?” などのように疑問詞や末尾の ? などで容易に疑問文であると同定できる事例もあれば, “The best way to kill ants.” や “Hotel Recommendations in San Diego” のように ? や疑問詞を含まないものが存在し、本研究ではこのような事例も要約とみなす。このような表現には多くのパターンが存在するため、質問文であるか否かを判定する網羅的な規則を記述するのが難しい。そこで、本研究では質問文を同定する網羅的な規則を設計するのではなく、タイトルの長さや質問テキストとタイトルの重複などの手がかりを用いてフィルタリングを行う方針を採用した。

すべての規則を適用後、251,420 対を獲得した。タイトルを要約とみなした場合の圧縮率は 0.18 である。これらの事例を抽出型および生成型手法の学習および評価に用いる。

5.1.2 データセットの妥当性検証

フィルタリング後のデータセットを先頭から 50 事例抽出し、分析すると 50 事例中 41 事例 (88%) が本研究で想定する質問テキスト-要約対とみなせる事例であった。質問テキスト-要約対とみなすことのできない 12% 程度 (9 事例) は以下に示す要因によりノイズとしてデータセットに含まれている。

- 質問タイトルから本文の語を照応する、またはその逆である (3 事例)
- タイトルもしくは本文のいずれかが質問ではない (3 事例)
- その他 (3 事例)

その他の要因には、タイトルと本文がフォーカスの異なる質問 (2 事例)、抽象的で質問内容が類推できないタイトル (1 事例) が存在する。

次に、人間が質問テキストを提示され作成する要約と、本研究で要約とみなすコミュニティ QA サイトに付与されたタイトルがどの程度近いものになっているか分析を行う。フィルタリング後のデータから 10 事例を抽出し、質問テキストを 10 人の作業者に提示し要約作成を依頼した。具体的には、クラウドソーシングサービスである Amazon Mechanical Turk⁴で、米国の高校および大学を卒業した作業者に依頼した。なお、作業者には要約とみなすタイトルの例をいくつか提示した上で、15 単語以内で要約を記述するよう指示した。

表 6 にクラウドソーシングによって作成した要約の実例と、本研究で要約とみなすタイトルを示す。例 1 は抽出型手法により要約可能な事例、例 2 は生成型手法が必要な事例である。例 1、例 2 ともにタイトルと作業者によって記述された要約のフォーカスは同等である。具体的には例 1 では、“薬物を用いずにアリを退治する方法”，例 2 では“中国で体調不良を避けるアドバイス”であるが、タイトルおよび人間による要約は同等のフォーカスを持つ質問文である。例 1 は 1 文目を抽出することで 16 単語以内の要約が作成可能な事例であるが、すべての作業者が言い換えを用いてより短い要約を作成した。正解タイトルも作業者の要約と同様に、質問テキスト中の “What are some good ways” を “Good ways” と単語除去により短い記述になっている。例 2 は体調不良の原因を食べ物だけにフォーカスする (質問テキストの前半部分の食べ物に関する記述を重視) か、旅行中を通して体調不良を起こさない方法にフォーカスする (後半の疑問文が食べ物以外の体調不良の要因も考慮すると考える) かによって、作成される要約が異なる。作業者 1、作業者 3、作業者 5 による要約およびタイトルはフォーカスを食べ物には限定しない要約となっている。

⁴ <https://www.mturk.com/>

表 6 クラウドソーシングにより作成した人間による要約とタイトルの比較

例 1 (抽出型手法で要約可能な事例)

質問テキスト:

What are some good ways to prevent/fight an ant invasion that don't involve poison?

I am using baits but they seem inadequate. Borax is totally ineffective.

タイトル:

Good ways to fight an ant invasion without poison.

クラウドソーシングによって作成した要約:

1. How do I get rid of ants without using poison?
2. Non-chemical ways to fight ants.
3. how to get rid of ants without using poison?
4. Any way to prevent ants that don't involve poison?
5. How can I combat an ant infestation without Borax or baits?

例 2 (生成的手法が必要な事例)

質問テキスト:

I'm planning a trip to China. I'm an adventurous eater and I'd like to think my stomach can handle anything that I encounter that appeals to me. But I know that's risky! Is there anything in particular precautions I should take or things I should avoid to stay healthy on the trip?

タイトル:

How can I avoid getting sick in China?

クラウドソーシングによって作成した要約:

作業者 1. Staying health during adventure in China.

作業者 2. when traveling to china what precautions should be taken before going and eating there?

作業者 3. Any precautions when traveling in China?

作業者 4. What are some food tips for staying healthy on my trip to China?

作業者 5. What are the risks of illness while traveling to China?

これらの例のように、人間が質問テキストから作成する要約およびタイトルはどちらも本文中の表現を言い換えることが多い。そのため、作業者間および作業者とタイトルを比較すると、完全に要約が一致するわけではない。しかし、そのような場合であっても多くの場合要約の質問文のフォーカスは同等となっている。

5.2 抽出型手法

抽出的な手法として、規則に基づく手法、機械学習に基づく手法を比較する。

5.2.1 規則に基づく手法

規則に基づく手法として、“リード文”、“リード疑問文”、“末尾疑問文”の3手法を比較する。リード文を抽出する手法は従来の抽出型要約研究においては強いベースラインとして知られている。リード文を抽出する手法は表4の例4において、“I want my chocolote...”から始まる文

を抽出する。質問要約課題においては出力も質問文になると考えられるため、規則を用いて同定した疑問文のうちもっとも先頭を抽出するリード疑問文、もっとも末尾に出現する疑問文を抽出する末尾疑問文とも比較を行う。リード疑問文は表4の例4において、“Why?”, 末尾疑問文は“What am I doing wrong?”となる。なお、疑問文の判定には以下の規則を用いる。

- 末尾の連続する記号に?を含む。
- 先頭単語が疑問詞、be 動詞、助動詞のいずれかである。

なお、規則を用いる手法であるリード疑問文、末尾疑問文は疑問文が入力中に出現しない場合、先頭文を出力する。フィルタリング後のデータセットにおいて、疑問文が存在しない文書の割合は37%であり、これらについてはすべての規則に基づく手法が先頭文を出力する。また、疑問文が1文以下である文書の割合は63%であり、このような場合にはリード疑問文と末尾疑問文の出力が等しくなる。

5.2.2 機械学習に基づく手法

機械学習に基づく手法としては、分類モデルに基づく手法、回帰モデルに基づく手法の2つを比較する。機械学習に基づく手法についても、質問文を優先的に出力するように設計する。

回帰モデルに基づく手法では、まず入力各文に対し ROUGE-2 F 値の予測値を出力する Support Vector Regression (SVR) (Basak, Pal, and Patranabis 2007) を学習する。学習済みの回帰モデルを用いて、入力各文に対し ROUGE-2 F 値を予測し、予測値のもっとも高い質問文を出力する。

分類に基づくモデルは、まず抽出した場合に ROUGE-2 F 値が最大になる疑問文を正例（要約に含めるべき文）、それ以外の文を負例（要約に含めるべきではない文）として Support Vector Machine (SVM) (Suykens and Vandewalle 1999) を学習する。学習した SVM を用い、入力各文を分類し正例と判定された文のうち先頭を出力する。SVM の出力は二値のラベルであり、回帰モデルのようにスコア最大となる疑問文に限定する後処理と組み合わせることができない。そこで、本研究では正例を疑問文に限定し、質問文を優先的に出力するような分類器が学習されるよう工夫する。なお、分類モデルはすべての文が負例と判定された場合には、先頭文を出力する。

SVR, SVM の学習には以下の素性を用いた。

- 単語 uni-gram
- 文長
- 分類対象文が1文目であるか
- 分類対象文が先頭の疑問文であるか
- 分類対象文の他に疑問文が存在するか

すべての素性は二値素性として表現した。また、単語 uni-gram 素性については訓練データ中

に5回以上出現する単語を用いた。文長素性については、文長が2単語以下, 5単語以下, 11単語以上, 15単語以上の4つの素性に分けて二値で表現した。

5.3 生成型手法

本研究では生成的な手法としてエンコーダ・デコーダモデル, 注意機構付きエンコーダ・デコーダモデル, コピー機構付きエンコーダ・デコーダを学習し, 比較を行う。

入力質問テキストは機械翻訳などの問題設定に比べ入力系列が長い。そのため, エンコーダ・デコーダモデルに加え, 注意機構付きエンコーダ・デコーダモデルを用い, 入力系列のうちデコード時に手がかりになる箇所に重みを付けながら要約を生成する。エンコーダ・デコーダモデルの学習においては低頻度語を UNK という特別なトークンに置き換え, 効率的に学習を行うという工夫がよく用いられる。そのため, 出力に UNK というトークンが含まれる状況がしばしば発生し, このような系列は要約としてそのまま提示することができない。質問要約においては, 入力質問テキストに出現する単語が出力にも含まれることが多い。そこで, 出力に入力質問テキストの単語を用いるコピー機構付きエンコーダ・デコーダモデルとの比較も行う。

エンコーダ・デコーダおよび注意機構付きエンコーダ・デコーダモデルには, Luong et al. (2015) の手法を用いる。コピー機構付きエンコーダ・デコーダには Gu et al. (2016) の手法を用いる。以下にこれらのモデルを簡単に説明する。

エンコーダ・デコーダモデル

エンコーダ・デコーダモデルは, エンコーダおよびデコーダという2つの要素から構成される。エンコーダは, 入力質問テキスト $\mathbf{x} = (x_1, \dots, x_n)$ から1単語ずつ受け取り, 連続値ベクトルによる内部状態 $\mathbf{h}_\tau$ に, Recurrent Neural Network (RNN) を用いて逐次変換する:

$$\mathbf{h}_\tau = f(x_\tau, \mathbf{h}_{\tau-1}). \quad (1)$$

f には Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997) や Gated Recurrent Unit (GRU) (Cho, van Merriënboer, Gulcehre, Bahdanau, Bougares, Schwenk, and Bengio 2014) などの関数を用いることができる。本研究では, 元論文の設定に従い, エンコーダ・デコーダモデルおよび注意機構付きエンコーダ・デコーダモデルでは LSTM を用い, コピー機構付きエンコーダ・デコーダモデルでは GRU を用いる。

デコーダは1つ前のタイムステップで生成された単語および内部状態を受け取り, 現在のタイムステップでの内部状態 $\mathbf{s}_t$ を計算する。計算した内部状態を用いて softmax 関数により単語 y_t の生起確率を計算できる。なお, デコーダの初期状態には入力質問テキスト中の単語をすべてエンコードし終えたときの最終状態 $\mathbf{h}_n$ を用いることにする:

$$\mathbf{s}_t = f(y_{t-1}, \mathbf{s}_{t-1}), \quad (2)$$

$$p(y_t|y_{<t}, \mathbf{x}) = \text{softmax}(g(\mathbf{s}_t)). \quad (3)$$

$\mathbf{x}$ が与えられた上での、出力要約 $\mathbf{y}$ が生起する条件付き確率は、以下のように出力単語の生起確率の積に分解される：

$$p(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^m p(y_t|y_{<t}, \mathbf{x}). \quad (4)$$

ただし、この確率の最大値を求めることは困難であるので、貪欲法により前から確率最大の単語を出力することで、要約を生成する。

学習時には訓練データにおける対数尤度を最大化するよう重みを更新する：

$$\log p(\mathbf{y}|\mathbf{x}) = \sum_{t=1}^m \log p(y_t|y_{<t}, \mathbf{x}). \quad (5)$$

注意機構付きエンコーダ・デコーダ

エンコーダ・デコーダでは、入力文書を逐次エンコードし、最終状態 $\mathbf{h}_n$ を文脈ベクトル $\mathbf{c}$ としてデコーダに渡す。そこで、注意機構付きエンコーダ・デコーダモデルでは、デコード時にエンコーダ側のどの単語に注意するかを考慮した重み付き文脈ベクトル $\mathbf{c}_t$ を考える：

$$\mathbf{c}_t = \sum_{\tau=1}^n \alpha_{t\tau} \mathbf{h}_\tau. \quad (6)$$

$\alpha_{t\tau}$ は入力質問テキストの t 番目の単語に与えられる重みで、以下のように softmax 関数を用いて計算される：

$$\alpha_{t\tau} = \frac{\exp(\mathbf{s}_t \cdot \mathbf{h}_\tau)}{\sum_{h'} \exp(\mathbf{s}_t \cdot \mathbf{h}')}. \quad (7)$$

入力側の重み付き文脈ベクトル $\mathbf{c}_t$ と $\mathbf{h}_t$ を用いて、入力単語への注意を考慮した内部状態 $\tilde{\mathbf{h}}$ を以下のように計算し、softmax 関数で確率値を出力する：

$$\tilde{\mathbf{h}} = \tanh(\mathbf{W}_c[\mathbf{c}_t; \mathbf{h}_t]), \quad (8)$$

$$p(y_t|y_{<t}, \mathbf{x}) = \text{softmax}(\mathbf{W}_s \tilde{\mathbf{h}}_t). \quad (9)$$

コピー機構付きエンコーダ・デコーダ

要約課題においては, 入力に含まれる単語が出力にも出現することが多い. そこで, 入力中の単語を出力により含めやすくするコピー機構を備えたデコーダを用い要約生成を試みる. Gu ら (2016) の提案したコピー機構付きのデコーダでは, 単語 y_t の生起確率を以下のように $score_{gen}$ と $score_{copy}$ の和として計算する:

$$p(y_t|y_{<t}, \mathbf{x}) = score_{gen}(y_t|\mathbf{s}_t, y_{t-1}, \mathbf{c}_t, \mathbf{x}) + score_{copy}(y_t|\mathbf{s}_t, y_{t-1}, \mathbf{c}_t, \mathbf{x}). \quad (10)$$

$score_{copy}$ は入力中の単語を出力に “コピー” するか否かをスコアリングする. 具体的には, y_t が入力文書中に含まれる ($y_t \in \mathbf{x}$) ならば, $score_{copy}$ は単語 y_t を要約に含める確信度をスコアリング関数 ϕ_c を用いて出力し, それ以外 ($y_t \notin \mathbf{x}$) の場合には 0 を出力する:

$$score_{copy}(y_t|\cdot) = \begin{cases} \frac{1}{Z} \sum_{j:x_j=y_t} \exp(\phi_c(x_j)) & (y_t \in \mathbf{x}) \\ 0 & (y_t \notin \mathbf{x}). \end{cases} \quad (11)$$

$score_{gen}$ は単語 y_t が語彙 V に含まれる場合に, その単語を要約に含めるか否かの確信度をスコアリング関数 ϕ_g を用いて出力する. 単語 y_t が語彙 V に含まれない場合には, 特別なトークン UNK を出力する確信度を出力する. ただし, 単語 y_t が入力文書 $\mathbf{x}$ に含まれる場合には 0 を出力する:

$$score_{gen}(y_t|\cdot) = \begin{cases} \frac{1}{Z} \exp(\phi_g(y_t)) & (y_t \in V) \\ \frac{1}{Z} \exp(\phi_g(UNK)) & (y_t \notin \mathbf{x} \quad \wedge \quad y_t \notin V) \\ 0 & (y_t \in \mathbf{x} \quad \wedge \quad y_t \notin V). \end{cases} \quad (12)$$

$score_{copy}$ は $y_t \in \mathbf{x}$ のときにスコアを出力し $y_t \notin \mathbf{x}$ のときに 0 になる. この機構により, 入力文書中に含まれる単語により高いスコアを付与し, 要約に含めやすくするようモデル化している. コピー機構付きのエンコーダ・デコーダでは $score_{gen}$ と $score_{copy}$ の和は $Z = \sum_{v \in V \cup \{UNK\}} \exp(\phi_g(v)) + \sum_{x \in \mathbf{x}} \exp(\phi_c(x))$ により総和が 1 になるように正規化されるため, 確率として扱うことができる. スコアリング関数 ϕ_c および ϕ_g の構築についての詳細は元論文 (Gu et al. 2016) を参照されたい.

6 実験

ROUGE-2 (Lin 2004) を用いた自動評価に加え, 人間の評価者による 5 段階評価を用いて, 各システムの性能を評価する.

6.1 実験設定

Yahoo! Answers Comprehensive Question and Answers version 1.0 をフィルタリングし獲得した 251,420 対を評価実験にも用いる。このデータのうち 90% を訓練データ、残りの 5% ずつをパラメータ調整用の開発データ、評価データにそれぞれ分割した。

分類モデルおよび回帰モデルで用いる SVM, SVR の実装には Liblinear (Fan, Chang, Hsieh, Wang, and Lin 2008) を用いた。カーネルには線形カーネルを用い、モデルパラメータ C は開発データでの ROUGE-2 F 値が最大となる値に設定した。なお、分類モデルの訓練データについては、訓練データ全体を使うと正例および負例の割合が不均衡になり学習が難しい。そこで、訓練データ全体から正例および負例をランダムに 10,000 事例ずつサンプリングし正例、負例の割合が 1:1 になるよう調整し分類器を学習した。

エンコーダ・デコーダの学習においては、単語埋め込みベクトルおよび隠れ層の次元を 256、バッチサイズは 64 に設定した。単語埋め込みベクトルは事前学習せず、他のモデルパラメータと同様に学習した。学習データに 1 回しか出現しない単語は UNK という特別なトークンで置換し、モデルパラメータ数を削減した。また、文末は EOS という特別なトークンで表現している。テスト時には開発データでの損失関数の値が最小となるモデルを用い、評価データでの性能評価を行った。エンコーダ・デコーダが 20 単語以内に EOS トークンを出力しない場合はデコードを中止し入力中の先頭の疑問文を出力する。入力中に疑問文が存在しない場合は、先頭文を出力する。EOS トークンが 20 語以内に出力されない現象はコピー機構付きエンコーダデコーダにおいて発生し、その割合は評価事例全体の約 20% である。

6.2 ROUGE による自動評価

本研究の課題設定においては、出力文数は 1 文という制約を課しているが文字数の制約はない。しかし、より短い文に適切な情報を埋め込んだ要約はより良い要約であると考えられる。そこで、本研究では ROUGE-2 の適合率による評価に加え ROUGE-2 F 値による評価を行う。

表 7 に各比較手法の ROUGE 値を示す。具体的には規則に基づく手法としてリード文、リード疑問文、末尾疑問文、機械学習に基づく手法として分類モデル、回帰モデル、エンコーダ・デコーダに基づく手法として注意機構付きエンコーダ・デコーダ、コピー機構付きエンコーダ・デコーダをそれぞれ示す。回帰モデルについては、文集合全体からスコアが最大になる文を選ぶモデルと、疑問文集合からスコア最大になる文を選ぶモデルの性能を示す。

規則に基づく手法において、リード疑問文を抽出する手法の ROUGE 値の適合率は 45.3 であり、末尾疑問文を抽出する手法の 42.6 やリード文を抽出する手法の 39.4 よりも良い性能を示した。F 値でも同様の順で良い性能を示した。先頭文を出力する規則は既存の要約課題においては強いベースラインとして知られる。しかし、質問要約課題においては先頭文を選択する手法より、リード疑問文や末尾疑問文など疑問文を優先的に出力するモデルのほうが良い性能を示

表 7 ROUGE-2 による性能評価

	適合率	F 値
リード文	39.4	27.0
末尾疑問文	42.6	33.9
リード疑問文	45.3	34.5
分類モデル	44.3	35.1
回帰モデル (疑問文から選択)	44.7	35.0
回帰モデル (文集合全体から選択)	44.7	29.7
エンコーダ・デコーダ	3.5	2.6
注意機構付きエンコーダ・デコーダ	38.5	38.5
コピー機構付きエンコーダ・デコーダ	47.4	42.2

す。また、リード疑問文の方が末尾疑問文よりも適合率において良い性能を示した。

分類モデルとリード疑問文を比較すると、ROUGE 値の適合率がそれぞれ 44.3 と 44.7 であり、ほぼ同等の性能を示した。ほとんどの入力にはただか 1-2 文の規則で同定可能な疑問文しか含まれず、分類器を用いてもリード疑問文よりも大きく性能を向上させることが難しいものと考えられる。回帰モデルおよび分類モデルの性能は、適合率による評価においてリード疑問文ベースラインよりもわずかに低くなった。このことから、リード疑問文ベースラインは質問要約課題において強いベースラインであることが分かる。また、機械学習に基づく手法についても規則に基づく手法と同様に、質問文を優先的に選択するモデルが良い性能を示した。

注意機構を持たないエンコーダ・デコーダモデルについては適合率で 3.5 と極端に低い性能を示した。Luong et al. (2015) が述べるように、注意機構を持たないエンコーダ・デコーダモデルは入力系列が長くなると性能が劣化する。本研究の入力は通常の機械翻訳の設定よりも長く、注意機構なしのエンコーダ・デコーダモデルでは正しくパラメータが学習されず、正しい要約を出力できなかった。注意機構付きエンコーダ・デコーダでは、この問題が解決され適合率および F 値が 38.5 と規則に基づく手法や機械学習に基づく手法よりも良い性能を示した。コピー機構付きエンコーダ・デコーダでは、適合率が 47.4、F 値が 42.2 とさらに性能が向上した。抽出的な手法はエンコーダ・デコーダを用いた生成的な手法よりも長い単語列を出力する傾向にある。そのため、適合率では良い性能を示すが、F 値においては生成型よりも劣る。コピー機構付きエンコーダ・デコーダでは適合率でも F 値でも抽出型よりも良い性能を示した。

6.3 人手評価

本研究では ROUGE による自動評価に加え、クラウドソーシングサービスである Crowdfunder⁵を用いた人手評価でも各モデルの性能を比較する。本研究では、英語として正

⁵ <http://crowdfunder.com>

しく、入力の問題テキストと同じ事柄について尋ねる要約がより望ましいと考える。そのため、人手評価においては、作業者に入力質問テキストと各モデルで生成した出力要約を提示し、“文法性”および“フォーカス”の2つの観点からより良い順に並べ替えるよう指示した。手法間の文法性やフォーカスに差異が見られない場合には、同順としても良いこととした。評価者には「例えばQAコミュニティサイトでのタイトルでの利用を想定する」という応用先を伝えた。“文法性”は英語として正しい文法で記述されているか、“フォーカス”は入力質問テキストと出力要約の尋ねている事柄が等しいかを表す評価基準である。比較手法としては、自動評価で良い性能を示した4つの手法を採用した。具体的には、人間による正解、規則を用いて抽出したリード疑問文、分類器が正例と判定した文のうち先頭、コピー機構付きエンコーダ・デコーダを用いた。評価には自動評価で用いた評価データから無作為に抽出した100事例を用い、各事例に対して3人の作業者が評価した。“文法性”、“フォーカス”の観点に基づく結果を表8、9にそれぞれ示す。この表では、行に示す各手法が列に示す手法よりもより良いと判定された回数を示している。例えば、人間によるタイトルはフォーカスの観点においてリード疑問文を抽出する手法よりも135回より良いと判定され、リード疑問文は人間によるタイトルよりも69回より良いと判定された。

人手評価では文法性、フォーカスどちらの観点においても自動評価と似た傾向を示した。具体的には、人間によるタイトル、コピー機構付きエンコーダ・デコーダモデル、リード疑問文および分類モデルの順でより良いと判定された。入力質問テキストに含まれる規則で同定可能な疑問文はたかだか1-2文であることが多く、リード疑問文と分類モデルはほぼ同様の出力をした。そのため、人手評価においても差が見られなかった。

コピー機構と人間によるタイトルを比較すると、フォーカス、文法性の観点においてそれぞ

表8 人手評価結果—文法性—

	人間によるタイトル	リード疑問文	分類モデル	コピー機構
人間によるタイトル	—	85	86	63
リード疑問文	51	—	10	54
分類モデル	54	10	—	53
コピー機構付きエンコーダ・デコーダ	69	79	82	—

表9 人手評価結果—フォーカス—

	人間によるタイトル	リード疑問文	分類モデル	コピー機構
人間によるタイトル	—	135	135	103
リード疑問文	69	—	11	72
分類モデル	70	10	—	68
コピー機構付きエンコーダ・デコーダ	89	107	103	—

表 10 各手法の出力要約

入力質問テキスト
The Simpsons is one of the funniest shows ever . Its one of my favorites . Do you like it ?
人間によるタイトル : Do you like The Simpsons?
リード疑問文/末尾疑問文 : Do you like it?
分類モデル : Do you like it?
注意機構付きエンコーダ・デコーダ : Do you like UNK?
コピー機構付きエンコーダ・デコーダ : Do you like The Simpsons?

れ 89 回, 69 回, コピー機構が人間のタイトルよりも良いと判定されている. このような事例の出力を分析すると, 人間の付与するタイトルには重要な単語まで除去しているものや, 完全な文ではない事例が含まれることがわかった. 例えば, コピー機構が “How do you stop the itching after shaving?” と出力する事例に対し, 人間のタイトルでは “after shaving” が除去され, さらに短い要約となっている事例などでは, コピー機構の方がより正しいフォーカスを持つと判定された. また, 人間のタイトルでは “The best way to get money?” など疑問詞を省略する事例がいくつか観測された. このような事例の文法性に基づく評価では, “What is the best way to get money?” など疑問詞を省略せずに出力する傾向のあるコピー機構よりも低く評価されている.

6.4 定性的分析

本節では, 実際の出力例を基に定性的な分析を行う. 表 10 に入力質問テキストと各モデルの出力例を示す. 抽出型手法であるリード疑問文や分類モデルでは, 照応詞の “it” が含まれ, フォーカスが不明確となることがある. このような照応の問題が, フォーカスでのスコアを下げる要因の一つであると考えられる. この例から単一質問文に要約する課題設定においては, 複数の文にまたがる情報をうまく組み合わせる必要があり, 抽出的手法では適切な要約を出力するのが難しいことが分かる. 生成型的手法である注意機構付きエンコーダ・デコーダモデルでは, “The Sompsons” のような低頻度語は特別なトークン UNK として出力され, ROUGE 値を下げる要因となっている. 同様の問題が機械翻訳課題でも報告されており (Bahdanau et al. 2014), この問題を解決するコピー機構付きエンコーダ・デコーダモデルでは, “The Simpsons” という重要語を正しく出力に含められている.

7 おわりに

本研究では新たな要約課題として, 複数文から構成される質問テキストを入力とし, その内容を端的に表現する単一質問文に要約する “質問要約” 課題を提案した. コミュニティ質問応

答サイトの投稿を用いた事例分析において、抽出型の手法では要約できない事例の存在を確認した。また、このデータをフィルタリングしたデータを用い、いくつかの抽出型および生成型要約モデルを構築し、比較した。先頭文を抽出する手法は既存の要約課題において強いベースラインとして知られるが、質問要約においては規則を用いて同定した疑問文のうち先頭を出力する手法より良い性能を示した。実験より、生成型の要約モデルが ROUGE-2 F 値においてより良い性能を示すことがわかった。構築した手法は人間によるタイトルよりも ROUGE 値および人手評価において、低い性能を示していることから、質問要約課題にはさらなる性能向上の余地が残されていると考える。

謝 辞

本稿は IJCNLP2017 に採録済みの論文を基にしたものです (Ishigaki, Takamura, and Okumura 2017). 本研究は JST さきがけ JPMJPR1655 の助成を受けたものです。

参考文献

- Bahdanau, D., Cho, K., and Bengio, Y. (2014). “Neural Machine Translation by Jointly Learning to Align and Translate.” *CoRR*. **abs/1409.0473**.
- Bank, M., Mittal, V. O., and Witbrock, M. J. (2010). “Headline Generation Based on Statistical Translation.” In *Proceedings of ACL2010*, pp. 318–325.
- Basak, D., Pal, S., and Patranabis, D. C. (2007). “Support Vector Regression.” *Neural Information Processing-Letters and Reviews*, **11** (10), pp. 203–224.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). “Learning Phrase Representations using Rnn Encoder-decoder for Statistical Machine Translation.” In *Proceedings of EMNLP2014*, pp. 1724–1734.
- Cohn, T. and Lapata, M. (2013). “An Abstractive Approach to Sentence Compression.” *ACM Transactions on Intelligent Systems and Technology (TIST)*, **4** (3), p. 41.
- Dorr, B., Zajic, D., and Schwartz, R. (2003). “Hedge trimmer: A Parse-and-trim Approach to Headline Generation.” In *Proceedings of HLT-NAACL2003 Text Summarization Workshop*, pp. 1–8.
- Duboue, P. A. (2012). “Extractive Email Thread Summarization: Can We Do Better Than He Said She Said?” In *Proceedings of INLG2012*, pp. 85–89.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). “LIBLINEAR: A Library for Large Linear Classification.” *Journal of Machine Learning Research*, **9**,

- pp. 1871–1874.
- Gu, J., Lu, Z., Li, H., and Li, V. O. (2016). “Incorporating Copying Mechanism in Sequence-to-Sequence Learning.” In *Proceedings of ACL2016*, pp. 1631–1640.
- Hirao, T., Isozaki, H., Maeda, E., and Matsumoto, Y. (2002). “Extracting Important Sentences with Support Vector Machines.” In *Proceedings of COLING2002*, pp. 1–7.
- Hochreiter, S. and Schmidhuber, J. (1997). “Long Short-Term Memory.” *Neural Computation*, **9** (8), pp. 1735–1780.
- Ishigaki, T., Takamura, H., and Okumura, M. (2017). “Summarizing Lengthy Questions.” In *Proceedings of IJCNLP2017*, Vol. 1, pp. 792–800.
- Kikuchi, Y., Neubig, G., Sasano, R., Takamura, H., and Okumura, M. (2016). “Controlling Output Length in Neural Encoder-Decoders.” In *Proceedings of EMNLP2016*, pp. 1328–1338.
- Li, C., Qian, X., and Liu, Y. (2013). “Using Supervised Bigram-based ILP for Extractive Summarization.” In *Proceedings of ACL2013*, pp. 1004–1013.
- Lin, C.-Y. (2004). “ROUGE: A Package for Automatic Evaluation of Summaries.” In *Proceedings of ACL2004 Workshop*, pp. 74–81.
- Luhn, H. P. (1958). “The Automatic Creation of Literature Abstracts.” *IBM Journal of Research and Development*, **2** (2), pp. 159–165.
- Luong, M.-T., Pham, H., and Manning, C. D. (2015). “Effective Approaches to Attention-based Neural Machine Translation.” In *Proceedings of EMNLP2015*, pp. 1412–1421.
- Mihalcea, R. and Tarau, P. (2004). “TextRank: Bridging Order into Texts.” In *Proceedings of EMNLP2004*, pp. 404–411.
- Oya, T. and Carenini, G. (2014). “Extractive Summarization and Dialogue Act Modeling on Email Threads: An Integrated Probabilistic Approach.” In *Proceedings of SIGDIAL2014*, pp. 133–140.
- Oya, T., Mehdad, Y., Carenini, G., and Ng, R. (2014). “A Template-based Abstractive Meeting Summarization: Leveraging Summary and Source Text Relationships.” In *Proceedings of INLG2014*, pp. 45–53.
- Peyrard, M. and Eckale-Kohler, J. (2016). “Optimizing an Approximation of ROUGE—a Problem-Reduction Approach to Extractive Multi-Document Summarization.” In *Proceedings of ACL2016*, pp. 1825–1836.
- Rush, A. M., Chopra, S., and Weston, J. (2015). “A Neural Attention Model for Sentence Summarization.” In *Proceedings of EMNLP2015*, pp. 379–389.
- Shen, D., Sun, J.-T., Li, H., Yang, Q., and Chen, Z. (2007). “Document Summarization using Conditional Random Fields.” In *Proceedings of IJCAI2007*, pp. 2862–2867.

- Suykens, J. A. and Vandewalle, J. (1999). “Least Squares Support Vector Machine Classifiers.” *Neural Processing Letters*, **9** (3), pp. 293–300.
- Tamura, A., Takamura, H., and Okumura, M. (2005). “Classification of Multiple-Sentence Questions.” In *Proceedings of IJCNLP2005*, pp. 426–437.
- Woodsend, K., Feng, Y., and Lapata, M. (2010). “Generation with Quasi-synchronous Grammar.” In *Proceedings of EMNLP2010*, pp. 513–523. Association for Computational Linguistics.
- Zajic, D., Dorr, B. J., and Schwartz, R. (2004). “Bbn/umd at duc-2004.” In *Proceedings of NAACL-HLT2004 Document Understanding Workshop*, pp. 112–119.

略歴

石垣 達也：2012 年茨城工業高等専門学校専攻科情報工学コース修了。2014 年東京工業大学総合理工学研究科博士前期課程修了。同年、同大学博士後期課程に進学。

高村 大也：1997 年東京大学工学部計数工学科卒業。2000 年同大学院工学系研究科計数工学専攻修了（1999 年はオーストリアウィーン工科大学にて研究）。2003 年奈良先端科学技術大学院大学情報科学研究科博士課程修了。博士（工学）。2003 年から 2010 年まで東京工業大学精密工学研究所助教。2006 年にはイリノイ大学にて客員研究員。2010 年から 2016 年まで同准教授。2017 年より同教授および産業技術総合研究所人工知能センター知識情報研究チーム研究チーム長。計算言語学、自然言語処理を専門とし、特に機械学習の応用に興味を持つ。

奥村 学：1962 年生。1984 年東京工業大学工学部情報工学科卒業。1989 年同大学院博士課程修了。同年、東京工業大学工学部情報工学科助手。1992 年北陸先端科学技術大学院大学情報科学研究科助教授。2000 年東京工業大学精密工学研究所助教授。2009 年同教授。現在に至る。工学博士。自然言語処理、知的情報提示技術、語学学習支援、テキスト評価分析、テキストマイニングに関する研究に従事。情報処理学会、電子情報通信学会、人工知能学会、AAAI、ACL、認知科学会、計量国語学会各会員。

(2018 年 8 月 1 日 受付)

(2018 年 11 月 1 日 再受付)

(2018 年 12 月 7 日 採録)