T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	ニューラルネットワークを用いた単一文書要約に関する研究
Title(English)	
著者(和文)	石垣達也
Author(English)	Tatsuya Ishigaki
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11264号, 授与年月日:2019年9月20日, 学位の種別:課程博士, 審査員:高村 大也,奥村 学,小野 功,中本 高道,長谷川 晶一
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11264号, Conferred date:2019/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Powered by T2R2 (Tokyo Institute Research Repository)

ニューラルネットワークを用いた
単一文書要約に関する研究

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
博士論文

指導教員: 教授　高村大也

石垣達也

2019 年 8 月

審査委員会

主査:　教授　高村大也

副査:　教授　奥村学

副査:　准教授　小野功

副査:　教授　中本高道

副査:　准教授　長谷川晶一

目 次

第 1 章　序論 **1**

1.1　研究背景 1

1.2　本研究の貢献 2

1.2.1　要約手法自体の高度化 3

1.2.2　要約対象テキストの多様化 3

1.3　本論文の構成 4

第 2 章　関連研究 **5**

2.1　要約モデルの発展とその評価法 5

2.1.1　抽出型手法 5

2.1.2　生成型手法 6

2.1.3　要約の評価 6

2.2　ニューラル抽出型要約モデルの定式化 7

2.2.1　単語エンコーダ 8

2.2.2　文エンコーダ 8

2.2.3　出力層 9

2.3　ニューラル生成型要約モデルの定式化 9

2.3.1　エンコーダ 9

2.3.2　デコーダ 10

2.4　談話構造 11

2.5　質問を対象とした研究 11

第 3 章　談話構造を考慮するニューラル抽出型要約モデル **13**

3.1　研究概要 13

3.2　定式化 15

3.3　リカレントニューラルネットワークに基づく抽出型要約器 16

3.3.1 単語エンコーダ 16
3.3.2 文エンコーダ 17
3.3.3 文スコアリング器 17
3.4 談話構造を考慮する注意機構 17
3.4.1 注意機構の動作概要 18
3.4.2 談話構造を考慮する注意機構の定式化 19
3.4.3 目的関数 21
3.5 実験 21
3.5.1 データ 21
3.5.2 評価手法 22
3.5.3 モデル学習 23
3.5.4 比較手法 23
3.6 結果および考察 24
3.6.1 DailyMail データセットでの結果 25
3.6.2 DUC2002 データセットでの結果 26
3.6.3 人手評価 27
3.6.4 定性的分析 27
3.7 談話構造を考慮する要約モデルのまとめ 28

第 4 章 質問を対象とする要約課題の提案とニューラル要約モデルの適用 30
4.1 研究概要 30
4.2 質問要約課題の特徴 31
4.3 質問応答サイトからの質問テキスト-要約対獲得と分析 33
4.3.1 分析 1: 質問の長さ 33
4.3.2 分析 2: 質問テキストとタイトルでの名詞の重複 34
4.3.3 分析 3: 抽出型 vs. 生成型 34
4.4 データと比較手法 35
4.4.1 データセット 35
4.4.2 抽出型手法 38
4.4.3 生成型手法 39
4.5 実験 42

4.5.1 実験設定 42
4.5.2 ROUGE 値による自動評価 43
4.5.3 人手評価 44
4.5.4 定性的分析 45
4.6 質問要約のまとめ 45

第 5 章 結論と今後の課題 **51**
5.1 結論 51
5.2 今後の課題 52

参考文献 **54**

謝辞 **62**

表 目 次

1 DailyMail データセットによる ROUGE 値による自動評価．DailyMail データセットで学習し，DailyMail データセットで評価．太字は最も良い性能を達成したモデルを示す．+記号はもっとも性能が近いベースライン手法と統計的有意な差を得たモデルである．#記号は開発セットでの ROUGE-2 が最大となったモデルである． 25
2 DUC2002 データセットでの ROUGE 値 26
3 ランダムに抽出した 100 事例を用いた人手評価 28
4 出力要約の例. 29
5 複数文質問とその要約 31
6 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 33
7 手法検討のための事例分析 33
8 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 47
9 Yahoo! Answers dataset 全体に含まれる質問テキストの長さ (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．) 47
10 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 48
11 ROUGE-2 による自動評価 49
12 人手評価結果 - 文法性 - 49
13 人手評価結果-フォーカス- 49
14 各手法の出力要約 50

図 目 次

1 本研究の貢献. 2
2 入力文書と談話依存構造の例. 14
3 談話構造を考慮する注意機構の動作概要. 18
4 学習された木構造の例 28

第1章　序論

1.1　研究背景

我々は日々，情報の取捨選択を行っている．例えば，新聞記事の見出しを見て本文を読むか否かを決定したり，論文のタイトルや要旨を手がかりに論文の全体を精読するか判断する．これらの見出し，タイトル，要旨は原文書の要点をまとめた “要約” と捉えることができる．インターネットが登場しテキスト情報が爆発的に増え続ける中，人々は大量の情報を効率的に処理する必要に迫られている．このような背景から，要約を計算機を用いて自動生成する技術が求められている．本研究では自動要約課題に取り組む．

本研究では自動要約課題のなかでも，とくに単一文書要約に着目する．単一文書要約は入力として 1 つの文書を想定する要約課題であり，自然言語処理の分野において半世紀以上に渡り取り組まれている課題の一つである (Luhn, 1958)．一方で，複数の文書を入力として想定する複数文要約課題 (McKeown & Radev, 1995) も存在する．単一文書要約課題では例えば，一つの新聞記事からその要旨や見出しを生成するといった応用先を想定するが，複数文書要約課題では事件の関連する新聞記事郡から事件をまとめた記事を生成するといった用途を想定する．入力および出力の特徴が異なることから，従来から異なる要約課題として扱われている．要約課題に多くの異なる設定が提案される中，従来から存在する単一文書要約の設定には依然として性能向上の余地がある．

単一文書要約課題の研究には二つの代表的な方向が存在すると考える．一つ目は要約手法自体の高度化である．この方向の研究では，研究者は新たな要約モデルを提案し，共通の評価用データを用いて性能を評価する．これにより，他の要約モデルとの比較が容易になる．上記の共通の評価セットには新聞記事や科学論文をもとにしたデータがしばしば用いられる．そのため，これらのデータセットで評価した要約モデルが新聞記事や科学論文とは異なる特徴を持つ会話テキストなどに対し同等の性能を得られるとは限らない．そこで，二つ目の方向として要約対象テキストの多様化が挙げられる．この方向の研究では，会話テキストや E メールなど新聞記事や科学論文とは異なる特徴を持つ要約対象に対して，独自のデータセットを作成し要約モデルの性能評価を行う．

単一文書要約課題の手法は抽出型手法と生成型手法に分類される．抽出型手法は原文書中

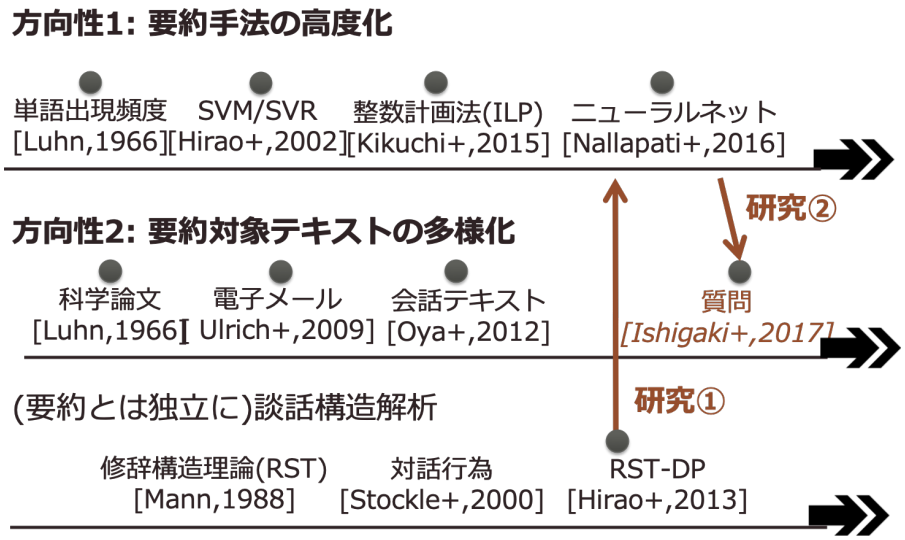


図 1: 本研究の貢献.

の文をそのまま抽出し要約として提示する手法である．生成型手法は，抽出するだけではなくさらに言い換えたり，圧縮することも許容される．すなわち，東京工業大学を東工大と略して要約に含めることも許容する．言い換えや圧縮が許容されることから，生成型手法はより短い字数の中により密な情報を埋め込むことが可能となり，原文書に対する要約の圧縮率の面では抽出型手法と比較し有利である．その一方，言い換えや圧縮に伴い非文を生成する可能性が高くなり実用上の安全性は相対的に低い．抽出型手法は文をそのまま抽出するため，非文を生成する可能性は低く実用上の安全性が高い．

1.2　本研究の貢献

本研究では図 1 に示すように，単一文書要約課題の研究における二つの方向に対して，ニューラルネットワークに基づく手法という軸をもとに貢献する．一つ目の方向に対しては，ニューラルネットワークに用いた既存抽出型要約手法に対し，文書中の文の意味的な関係である談話構造を考慮する新たな機構を導入した新たな要約モデルを提案する．二つ目の方向に対しては，質問を対象とする新たな要約課題を提案し，ニューラルネットワークに基づく要約モデルが特に良い性能を示すことを報告する．以下にそれぞれの方向と本研究の貢献について詳述する．

1.2.1 要約手法自体の高度化

前述したように単一文書要約のための手法は単語の出現頻度などを手がかりとして，規則やヒューリスティックに基づく手法から始まり，その後機械学習の適応，整数計画法によって制約付き最適化問題として文選択を行う手法を経て，2016 年以降のニューラルネットワークに基づく手法へ発展し現在に至る．ニューラルネットワークに基づく手法は生成型もしくは，生成型と抽出型を組み合わせるようなモデルが活発に提案されているが，抽出型手法にも依然として改善の余地がある (Hirao, Nishino, & Nagata, 2017)．本研究では 1 つ目の方向として，ニューラルネットワークに基づく抽出型要約モデルに対し文間の談話構造を考慮する機構を新たに提案する．2016 年以降は Nallapati ら (Nallapati, Zhai, & Zhou, 2017) のモデルを皮切りに，多くのニューラルネットワークを用いる要約モデルが提案されている．本研究でも彼らの手法を拡張する．彼らの手法は原文書を文の系列と捉え，ニューラルネットワークを用いて原文書中の各文をベクトルに変換する．このベクトルを重要度を計算するための出力層と呼ばれるニューラルネットワークに入力し，重要度スコアを得る．一方で，ニューラルネットワークが要約課題に適用される以前の整数計画法に基づく手法では，原文書をただ文の系列と捉えるのではなく，文間の意味的な関係である談話構造に着目し木構造として表現することで，性能が向上することが報告されている．本研究ではニューラルネットワークに基づく抽出型要約手法においても，原文書を文の系列としてだけでなく，木構造として捉える機構を導入する．提案モデルを新聞記事を元にした共通のベンチマークデータセットである DailyMail データセットを用い評価する．

1.2.2 要約対象テキストの多様化

一つ目の方向である要約手法自体の高度化を目指す研究においては，新聞記事や科学論文をもとにした共通のベンチマークデータセットを用いて要約モデルの性能評価を行う．しかし，新聞記事や科学論文とは異なる特徴を持ったテキストが存在する．例えば，E メールスレッドは新聞記事や科学論文とは異なる文体で書かれる．具体的には，これらのテキストは話者が 2 人以上となり，文書中に引用符や質問文などが含まれる．このように，新聞記事や科学論文をもとにした共通データセットでは要約モデルの性能が正しく評価できない種類のテキストが存在する．これまでには，E メールスレッドや会話テキスト (Duboue, 2012; Oya & Carenini, 2014; Oya, Mehdad, Carenini, & Ng, 2014) を対象とする要約課題が提案されており，独自のデータセットを用いて評価されてきた．本研究では，質問テキストに着目した要約課題を新たに提案する．質問テキストは，インターネット上のコミュニティ質問応答

サイトの登場により，その量が増えている．質問テキストを明示的に要約対象として扱った研究は少なく，そもそも抽出型手法で要約可能であるのか，生成型の手法が必要なのかが明らかではない．そこで，本研究ではコミュニティ質問応答サイト上の質問投稿を対象に必要な要約手法についての分析を行い，考えうる様々な抽出型および生成型要約手法を実際に質問テキストに適用し比較した．その結果，生成型要約手法が必要となる事例の存在を確認し，ニューラルネットワークに基づく生成型要約手法が質問要約課題において良い性能を示すことを確認した．これらの知見は要約研究の二つ目の方向に対し貢献するものである．

1.3　本論文の構成

本章以降の構成について述べる．2 章では，ニューラルネットワークを用いた既存要約モデルおよび 1 つ目の研究に関連する談話構造解析課題や質問を対象とした既存研究について説明する．3 章では，1 つ目の研究であるニューラル抽出型要約モデルに談話構造に関する情報を考慮する機構を導入する研究について説明する．この章では抽出型要約器について定式化したのち，談話構造を考慮する機構およびニューラルネットワークの学習手法について述べる．最後に談話構造を考慮した実験により，性能が向上することを示す．4 章では，質問を対象とする新たな要約課題の提案を行う．この章では，まずはじめにコミュニティ質問応答サイトである Yahoo! Answers から質問とその要約の対を獲得しデータセットを作成する手法について議論する．その後，作成したデータセットの分析および抽出型および生成型要約モデルの比較実験について述べる．5 章では全体のまとめを行い，今後の課題について議論する．

第2章　関連研究

本章では本研究に関連する既存研究について議論する．はじめに，要約モデルの発展について抽出型手法および生成型手法に分けて概説する．また，要約モデルをどのように評価するかという点についても述べる．その上で，本研究で対象とするニューラルネットワークに基づく要約モデルの基本的な仕組みについて説明する．抽出型モデルおよび生成型モデルにおいてやや異なるネットワーク構造を持つことから，本章ではニューラルネットワークに基づく抽出型要約モデルについてまず説明したのち，生成型要約モデルについて説明する．次に，一つ目の貢献として述べた談話構造を考慮する抽出型要約モデルに関連する研究として，談話構造解析に関する研究を取り上げる．談話構造解析に関する研究自体は，要約研究とは独立して行われている．最後に，二つ目の貢献として述べた質問を対象とする要約課題に関連する研究について述べる．

2.1　要約モデルの発展とその評価法

要約モデルの発展について，抽出型手法および生成型手法に分けて述べる．その後，要約モデルを評価する手法について説明する．

2.1.1　抽出型手法

古典的な抽出型手法としては，Luhn ら (Luhn, 1958) の単語の出現頻度を手がかりに文の重要度を決定する手法が挙げられる．文の重要度を決定する手法はその後広く研究され，要約と原文書の意味的な類似度が近くなるよう文を抽出する手法 (Dorr, Zajic, & Schwartz, 2003; Kobayashi, Noguchi, & Yatsuka, 2015) や要約に含めるべき文を二値分類問題として定式化する教師あり学習に基づく手法 (Hirao, Isozaki, Maeda, & Matsumoto, 2002; Shen, Sun, Li, Yang, & Chen, 2007)，回帰モデルを用いて重要度を決定する手法 (Peyrard & Eckale-Kohler, 2016; Li, Qian, & Liu, 2013) などが存在する．グラフに基づき文の重要度を計算する手法も存在し，LexRank(Erkan & Radev, 2004) はシンプルかつ良い性能を示すことが知られている．上述の文の重要度計算に関する研究に加え，2007 年以降には，整数計画法として文選択問題を定式化する手法が登場した．これらの手法では，出力長制約や文間の意味的な関係（談話構造）などに基づく制約を満たしながら文の重要度の総和が最大になる文を探索する (Kikuchi,

Hirao, Takamura, Okumura, & Nagata, 2014)．2016 年以降にはニューラルネットワークを用いて文の重要度を計算する手法 (Cheng & Lapata, 2016; Nallapati et al., 2017) が提案された．ニューラルネットワークに基づく手法は現在活発に研究されており，本研究でもこれに着目する．

ニューラルネットワークを用いた要約モデルとしては，生成型もしくは生成型と抽出型を組み合わせるような手法に研究の主軸がシフトしているが，Hirao ら (Hirao et al., 2017) が指摘するように，抽出型要約モデルそのものにも依然として性能向上の余地が残されている．よって，抽出型手法による単一文書要約の研究も依然として盛んに行われている (Cheng & Lapata, 2016; Nallapati et al., 2017)．

2.1.2　生成型手法

生成型手法としては，グラフに基づく手法 (Parveen, Ramsl, & Strube, 2015; Wan, 2010) などが存在するが，とくにニューラルネットワークに基づく生成型手法が良い性能を示しており，近年活発に研究されている (Nallapati, Zhou, dos Santos, glar Gulehrea, & Xiang, 2016; See, Liu, & Manning, 2017; Gu, Lu, Li, & Li, 2016)．ニューラルネットワークに基づく手法では，エンコーダデコーダと呼ばれるネットワーク構造がしばしば用いられる．エンコーダデコーダは，入力文書を読み込むエンコーダと呼ばれるネットワークと，要約に含める単語を逐次出力するデコーダと呼ばれるネットワークに分けて構築される．エンコーダデコーダはもともと機械翻訳課題向けに提案 (Luong, Pham, & Manning, 2015; Bahdanau, Cho, & Bengio, 2015) されたが，2015 年に文圧縮課題 (Filippova, Alfonseca, Colmenares, Kaiser, & Vinyals, 2015) および見出し生成課題 (Rush, Chopra, & Weston, 2015) に適用され，文要約課題への有効性が示された．2016 年には生成型文書要約課題にエンコーダデコーダを適用する研究 (Nallapati et al., 2016) が登場し，文書要約課題においても良い性能を示すことが明らかになった．その後，原文書に含まれる単語をより要約にも含めやすくする工夫を加えたコピー機構付きエンコーダデコーダ (See et al., 2017; Gu et al., 2016) などの拡張が行われ，さらなる性能向上が達成されている．

2.1.3　要約の評価

要約の評価には自動評価指標および人手評価の 2 種類が良く用いられる．

要約課題では，同一の入力文書に対し作成者によって異なる要約を作成する場合がある．すなわち，要約には多くの場合複数の正解が存在する．こうした状況から，要約モデルの評価には共通の自動評価指標である ROUGE (Lin, 2004) が良く用いられる．ROUGE 値の計算

においては，まずシステムの出力と正解となる要約（参照要約）を用意する．参照要約は例えば，人間の手作業によって記述された要約などになる．一つの文書に対する参照要約は複数存在しても良い．ROUGE は，人間の要約に含まれる n-gram をどの程度システムによる要約が再現できたかを評価する:

$$ROUG-_n \quad = \quad \frac{\text{システム要約中の } n-gram \ \cap \text{参照要約中の } n-gram}{\text{参照要約中の } n-gram} \tag{1}$$

ROUGE-1 は参照要約のユニグラムをシステム出力がどの程度の割合で出力できたかを示し，ROUGE-2 はバイグラムの再現率を表す．ROUGE-2 はとくに，情報量の観点において人間による評価と相関を持つことが知られている．また，ROUGE-L は最大の連続する単語系列を考慮しながらスコアを計算する．また，ROUGE-F は異なる出力長でも比較ができるよう長さによる正規化を行う．他にもいくつかの亜種が存在するが，この 4 つの指標がよく用いられる．

ROUGE による自動評価では，要約に含まれる文の意味的なつながりの良さ (一貫性) や読みやすさ (Readability) を適切に評価することが出来ない．そのため，多くの既存研究では ROUGE による自動評価を基本として，人間が目視で要約を確認する人手評価を組み合わせ用いる．例えば，DUC の quality question という要約の評価基準をまとめた Web ページ[1]では,「文性」,「冗長性」,「参照の明確さ」,「焦点」および「構造と一貫性」のについて 1 点から 5 点で評価する手法を説明している．本研究でも，ROUGE による自動評価に加え人手評価を用いる．

2.2 ニューラル抽出型要約モデルの定式化

ニューラルネットワークに基づく抽出型要約器について，その仕組みを述べる．抽出型要約器はエンコーダと出力層という 2 つのネットワークから構成される．エンコーダは入力文書の各文を読み込み，各文に対し 1 つずつ固定長ベクトルを生成する．出力層では，エンコーダの生成した文を表現するベクトルを，当該文を抽出するか否かを表す確率分布に変換する．各文に対し抽出するか否かを表す確率分布が得られると，最終的には確率値の大きな文から順に出力長制約に達するまで要約に含める．以下にエンコーダ，出力層のネットワークについて定式化する．

[1] https://duc.nist.gov/duc2007/quality-questions.txt

2.2.1　単語エンコーダ

ニューラルネットワークを用いた抽出型要約器は，まずエンコーダと呼ばれるネットワークを用いて入力文書中の各文をベクトルに変換する．ベクトルへの変換は単語エンコーダおよび文エンコーダという 2 つのネットワークが階層的に接続された構造を持っている．

単語エンコーダは入力文書 d の文 x_i 中の単語 $w_{i,j}$ 対して，リカレントニューラルネットワークを用いて文書中の各文に対し，単語を先頭から逐次読み込み文 x_i を表現するベクトル h_i を得る．具体的には，文 x_i 中の単語系列 $\{w_{i,1}, w_{i,2}, ..., w_{i,j}, ..., w_{i,M}\}$ に対し両方向 Long Short Term Memory (Bi-LSTM) を用いて構成されるリカレントニューラルネットワークを用いて以下のように h_i を得る．

$$\overrightarrow{h}_{i,j} = LSTM(\overrightarrow{h}_{i,j-1}, emb(w_{i,j})), \tag{2}$$

$$\overleftarrow{h}_{i,j} = LSTM(\overleftarrow{h}_{i,j+1}, emb(w_{i,j})) \tag{3}$$

また，$emb(\dot{})$ は単語 $w_{i,j}$ の埋め込みベクトルを表す．単語を埋め込みベクトルに変換する手法はいくつか存在するが，本研究では word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) を用いる．

単語エンコーダは最終的に文 x_i に対し，前向きおよび後ろ向きの LSTM の出力を以下のように結合したベクトル h_i $(1 \leq i \leq N)$ を文 x_i を表現するベクトルと考える．

$$h_i = [\overrightarrow{h}_{i,M_i}; \overleftarrow{h}_{i,1}]. \tag{4}$$

2.2.2　文エンコーダ

単語エンコーダによって生成されたベクトル h_i は文の情報を保持している．抽出型要約器においては，文 x_i そのものの情報に加え周辺に出現する文の情報も考慮しながら，相対的な文 x_i の重要度を計算したい．そこで，ベクトル h_i をさらに別の両方向 LSTM によって読み込む階層的文エンコーダという仕組みがしばしば用いられる (Nallapati et al., 2017; Cheng & Lapata, 2016)．具体的にはおよび前向きおよび後ろ向き LSTM のを以下のように結合し，文脈を考慮したベクトル H_i を各文 x_i に対し生成する．

$$\overrightarrow{H}_i = LSTM\ (\overrightarrow{H}_{i-1}, h_i), \tag{5}$$

$$\overleftarrow{H}_i = LSTM(\overleftarrow{H}_{i+1}, h_i), \tag{6}$$

$$H_i = [\overrightarrow{H}_i; \overleftarrow{H}_i]. \tag{7}$$

単語エンコーダと文エンコーダを階層的に接続し得られたベクトル H_i は，文 x_i とその周辺に出現する文の文脈も含めた情報が格納されたベクトルと考えられる．よって，H_i を用い

て周辺の文脈を考慮しながら文 x_i を要約に含めるべき確率を計算することができる．確率計算は出力層が行う．

2.2.3 出力層

出力層は上述の文エンコーダで得られたベクトル H_i を入力として受け取り，文 x_i を要約に含めるか否かの確率分布を出力する:

$$p(y_i|\mathbf{x}, \theta) = softmax(\mathbf{W}_o H_i). \tag{8}$$

この式において，出力層は文を表現するベクトル H_i を重み行列 $\mathbf{W_o} \in \mathcal{R}^{|\mathbf{H_i}| \times \mathbf{2}}$ をかけあわせて 2 次元ベクトルに圧縮する．2 次元ベクトルのそれぞれの次元は文 x_i を抽出する確信度および抽出しない確信度に対応する．これらの確信度は確率値として正規化されておらず，そのままでは扱いづらいためソフトマックス関数を用いて，確率値に変換する．すべての文に対し，抽出するか否かを表現する確率分布が求まれば，出力長制約を満たす範囲で抽出する確率値の大きな順から文を抽出し要約に含めることで，要約が出力できる．

上記がニューラルネットワークに基づく抽出型要約モデルの単純な定式化である．近年発表されているモデルでは，このモデルを基本とし，出力層の中で原文書と対象文の類似度などを考慮する機構が提案されている (Nallapati et al., 2017)．また，出力層においても LSTM を用いる亜種も存在するが第 3 章において詳述する．

2.3 ニューラル生成型要約モデルの定式化

ニューラルネットワークによる生成型要約器はエンコーダとデコーダという 2 つのリカレントニューラルネットワークから構成される．エンコーダでは，抽出型モデルと同様に入力文書の各文を LSTM によって読み込み，文書全体を単一のベクトルで表現する．デコーダでは，エンコーダの出力するベクトルを受け取り，別のリカレントニューラルネットワークを用いて，事前定義した辞書中の単語についてどの単語を出力するか確率分布を出力する．確率値がより高くなるような単語系列を要約とみなし出力する．以下にエンコーダおよびデコーダについて順に説明する．

2.3.1 エンコーダ

エンコーダは入力文書を単語の系列とみなし，文書をベクトル表現に変換する．リカレントニューラルネットワークを用いて，文書中の単語を逐次読み込む．以下のように単語の埋め込みベクトル及び直前の LSTM の出力を入力として，次の状態ベクトルを出力するという

ステップを入力文書すべての単語について繰り返す:

$$\overrightarrow{h}_i = LSTM(\overrightarrow{h}_{i-1}, emb(w_i)), \tag{9}$$

$$\overleftarrow{h}_i = LSTM(\overleftarrow{h}_{i+1}, emb(w_i)) \tag{10}$$

前向き LSTM は単語を先頭から読み込んでいき，後ろ向き LSTM は単語を末尾から読み込む．前向き LSTM および後ろ向き LSTM の最終状態ベクトルは，単一のベクトル h に結合されデコーダに引き渡される:

$$h = [\overrightarrow{h}_{i,M}; \overleftarrow{h}_{i,1}]. \tag{11}$$

ここで，M は入力文書中の単語数である．

2.3.2　デコーダ

デコーダは 1 つ前のタイムステップで生成された単語の埋め込み表現および LSTM の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 $\boldsymbol{s}_t$ を計算する．計算した内部状態を用いてソフトマックス関数により辞書中の単語 y_t に対し要約に含めるべき確率を計算する:

$$\boldsymbol{s}_t = LSTM(y_{t-1}, s_{t-1}), \tag{12}$$

$$p(y_t|y_{<t}, \boldsymbol{x}) = softmax(g(\boldsymbol{s}_t)). \tag{13}$$

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 $\boldsymbol{h}$ を用いることにする:

入力文書 $\boldsymbol{x}$ が与えられた上での，出力要約 $\boldsymbol{y}$ 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:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{m} p(y_t|y_{<t}, \boldsymbol{x}). \tag{14}$$

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したり，ビームサーチを用いてより滝確率値を得られる系列を探索する．

重み行列のパラメータを学習する際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:

$$\log p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{t=1}^{m} \log p(y_t|y_{<t}, \boldsymbol{x}). \tag{15}$$

2.4 談話構造

次に一つ目の方向に関する研究に関連する談話構造解析について説明する．談話構造は入力文書の単位の意味的な構造のことであり，修辞構造理論 (Mann & Thompson, 1988) や Penn Discourse Tree Bank (Prasad, Dinesh, Lee, Miltsakaki, Robaldo, Joshi, & Webber, 2008) によるアノテーション基準をを用いてしばしば表現される．とくに修辞構造理論は文書全体を木構造で表現できることから，要約課題との相性が良い．この理論において，文は Elementary Discourse Unit と呼ばれるおおよそ節に対応する粒度の単位の組み合わせにより表現され，文書は文や段落の組み合わせで表現されると仮定される．要約課題においては，文を抽出する設定が多く用いられるため本研究では文を文書の最小の構成要素と考える．この理論において，隣接する文は，補足や対比といったようになにかしらの意味的なつながりを持っていると仮定する．補足される文はより重要と考えられることから核と呼び，補足する文は衛星と呼ぶ．補足関係のように核と衛星が決まる関係もあれば，対比関係のように相対的な重要度が決められずどちらの文も核とする関係の双方が存在する．隣接する文を 1 つの単位に結合し，さらに隣接する単位との関係を同定するという手順を繰り返すことで，入力文書を句構造構造木として表現する．

談話構造解析は自動要約とは独立して研究されている．しかしながら，談話構造に関する情報は要約する際の重要な手がかりととして有効であることが示されている．例えば，Hirao ら (Hirao, Yoshida, Nishino, Yasuda, & Nagata, 2013) はこれらを依存構造で表現する手法を提案した．依存構造で表現することにより，どの文がより核となる文であるか容易に同定できるようになる．核となる文により高いスコアを与え，談話構造木を保持するように文を抽出することで整数計画法に基づく要約器の性能が向上することが報告されている (Hirao et al., 2013; Kikuchi et al., 2014)．本研究ではこのような知見を背景に，ニューラルネットワークに基づく抽出要約器を談話構造を手がかりとした情報を考慮するよう拡張する．これにより，要約研究の 1 つ目の方向である要約モデルそのものの高度化に対する貢献を目指す．

2.5 質問を対象とした研究

次に二つ目の研究に関連する研究として，質問を対象とした自然言語処理に関する研究について説明する．自然言語処理分野において，質問を扱う研究課題には質問生成や質問応答システムなどいくつか存在する．質問生成課題においては，入力は質問文ではない叙述文，出力はその文が答えとなる質問文であり，主に教材の自動生成などの応用先が想定されている．要約課題にもっとも関連するのは質問応答システムの研究である．多くの質問応答システムで

は，単一質問文を入力として受け取りなにかしらの答えを出力する．これらのシステムでは，入力として単一文質問を想定するため複数文から構成される質問を入力とすることができない．そこで，Tamura ら (Tamura, Takamura, & Okumura, 2005) は複数文から構成される質問からもっとも核となる 1 文を抽出し，質問応答システムに入力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示した．質問タイプ同定は質問応答システムにおいて重要な前処理ステップであり，この処理での性能向上は質問応答システム全体の性能を向上させると考えられる．Tamura ら (Tamura et al., 2005) の手法は複数文質問を単一文質問に要約するという要約課題とも捉えることができる．Tamura らの研究では質問タイプ同定器の性能を向上させることに主眼が置かれている．本研究の 2 つ目の貢献として，ユーザに直接提示し理解可能な質問を要約として出力する質問要約課題を新たに提案する．これにより，要約研究の 2 つ目の方向である要約対象テキストの多様化に対しての貢献を目指す．

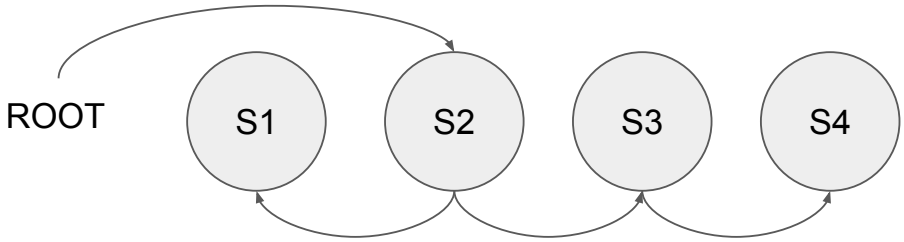
第3章　談話構造を考慮するニューラル抽出型要約モデル

本章では，要約研究の一つ目の方向に対する貢献に関する研究について述べる．

3.1　研究概要

抽出型単一文書要約課題においては，これまで多くの手法が提案されている．リカレントニューラルネットワークに基づく要約モデル (Cheng & Lapata, 2016; Nallapati et al., 2017) が 2016 年以降，非常に良い性能を示している．しかしながら，既存のリカレントニューラルネットワークに基づく手法は入力文書を単語および文の系列とみなす．そのため，既存手法では本来文書に内在する文間の意味的なつながり（談話構造）が無視されている．我々の知る限り，Cohan ら (Cohan, Dernoncourt, Kim, Bui, Kim, Chang, & Goharian, 2018) の取り組んだ科学論文の要約課題の研究が唯一，元文書内の表現の意味的な情報を考慮するモデルである．彼らの手法では，入力の論文を文の系列ではなく “Introduction” や “Related Work” などの節の系列とみなすことで，要約モデルの性能が向上することが報告されている．彼らの手法でも “Introduction” や “Related Work” といった節に関する意味的な情報を利用しつつも，入力文書は系列とみなされている．そのため，木構造で表現された談話構造をリカレントニューラルネットワークに基づく文書要約器に組み込むことで，要約器の性能が向上するか否かは明らかではない．

談話構造とは，文間の意味的な関係を表現する枠組みである．談話構造を表現する理論はいくつか提案されているが，Rhetorical structure theory (RST) (Mann & Thompson, 1988) が文書を木構造として表現するため，要約課題との相性が良い．ゆえに，もっともよく利用されている．RST において文書は文などの単位の集合と考えられ，文間の関係が木構造で表現される．RST によって表現された木構造の例を図 2 に示す．この図において，ノードは文に対応し，依存構造木によって文書が木構造で表現される．具体的な文間の関係に着目すると，$S2$ は $S1$ によってもたらされた結果に言及しており，この 2 つの文は “因果” の関係にある．因果関係においては結果に言及する文のほうがより重要であると考えられるため，依存構造木においては $S2$ を親とし，$S1$ を子とするエッジで結ばれる．さらに，$S3$ および $S4$ は



S1: In Afghanistan, many girls are forbidden from riding bicycles, but they are allowed to ride skateboards.
S2: So when Australian skateboarder Oliver Percovich created a non-profit organization that aims to connect youth in Afghanistan with education through skateboarding, it shouldn't have been a surprise that 40 percent of the program's students would be girls.
S3: In 2012, photographer Jessica Fulford-dobson decided to take pictures of these skater girls.
S4: Scroll down for video.

図 2: 入力文書と談話依存構造の例.

$S2$ の言及に対し補足的な情報を加えており，RST での関係定義においては “補足” 関係にある．補足される文のほうがより重要であるので，$S2$ を親，$S3$ および $S4$ を子とするノードで結ばれる．このような木構造を正しく同定することができれば，より根に近い位置にある文がより重要であると考えられ，文の相対的な重要度のより正確な計算に役立つ．

ニューラルネットワークを用いない要約モデルにおいては，木構造で表現された談話構造を組み込むことで性能が向上することが報告されている．例えば，整数計画法に基づく要約モデルの文選択処理において，要約を木構造を保持するように文を選択することで自動評価指標である ROUGE 値や人間による評価が向上することが報告されている (Hirao et al., 2013; Kikuchi et al., 2014)..これらの手法では元文書を既存の談話構造解析器で解析し，解析した情報を要約器で用いる．したがって，要約器の性能は談話構造解析器の性能に大きく依存する．談話構造解析器の性能は文書の長さが長くなったり，解析器を学習したドメイン以外の文書に適用すると性能が著しく劣化することが知られている．

本研究では談話構造解析器の解析誤りの影響を抑えつつ，近年良い性能が報告されているリカレントニューラルネットワークに基づく手法の利点を活用するために，元文書の談話構造と文の重要度のスコアリング器を同時に学習する新たな要約モデルを提案する．提案モデルでは，リカレントニューラルネットワークの注意機構により元文書の談話構造を表現し，文の重要度をソフトマックス関数を用いて構築した出力層で表現する．Kamigaito ら (Kamigaito,

Hayashi, Hirao, & Nagata, 2018) は同様の機構を文圧縮課題に対して提案し，性能が向上することを報告している．彼らの手法では単語間の依存関係を注意機構を用いて表現し，単語を要約に含めるか否かの確率値をソフトマックス関数で構築された出力層で表現する．本研究では，この手法を文書レベルの要約課題に拡張する．文書要約への拡張においては，文間の依存関係を注意機構を用いて表現し，文を要約に含めるか否かの確率値を出力層で表現する．

提案モデルの性能を ROUGE(Lin, 2004) を用いていくつかのベースライン手法と比較した．その結果，提案モデルが強い比較手法と同等もしくはより良い性能を示した．

3.2 定式化

本研究では抽出型文書要約を文抽出問題として定式化する．本節では本研究で用いる表記について説明し，文抽出問題を定式化する．

入力文書 $\boldsymbol{x}$ は文の系列 $x_1, ..., x_N$ として表現される．さらに，文 x_i は単語系列 $w_{i,j}$ $(1 \leq j \leq M_i)$ で表現される．また，x_0 は依存構造木における擬似的な根を表現する． 要約器は入力文書に対し，文を抽出するか否かを表現する二値系列 $\boldsymbol{y} = y_1, ..., y_N$ を出力する．ここで，$y_i = 1$ ならば文 x_i を抽出し要約に含め，$y_i = 0$ ならば x_i を抽出しない．文を抽出するか否かを表現する二値系列 $\boldsymbol{y}$ はニューラルネットワークによって計算される確率分布 $p(y_i|\boldsymbol{x}, \theta)$ により決定される．ここで，θ は学習されたパラメータである．本研究では $p(y_i|\boldsymbol{x}, \theta)$ の総和が最大となる二値系列を $\boldsymbol{y}$，貪欲法により探索する．

したがって，本研究では抽出すべき文に対しより高い確率値を出力し，抽出すべきでない文に対しては低い確率値を出力するようパラメータ θ をデータセット D から学習することである．データセット D 内の各事例は $(\boldsymbol{x}, \boldsymbol{E}, \boldsymbol{y})$ の 3 要素からなる．ここで，$\boldsymbol{E}$ は文書の依存構造を表現する隣接行列である．例えば，隣接行列中の要素 $E_{k,l}$ は文 x_k から x_l へのエッジが存在する場合に 1 をとる．よって，図 2 に示す文書に対しては以下のような隣接行列ができる．

$$\boldsymbol{E} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{16}$$

3.3　リカレントニューラルネットワークに基づく抽出型要約器

本研究ではリカレントニューラルネットワークに基づく要約モデルに，入力文書の談話構造を考慮する機構を導入する．本節では，まず基本となるリカレントニューラルネットワークに基づく要約器について説明し，その後談話構造を考慮する拡張について述べる．

基本となるモデルは 2 つの要素から構成される．1 つ目は階層的文書エンコーダ，2 つ目は注意機構付き文スコアリング器である．階層的文書エンコーダはさらに単語エンコーダ，文エンコーダの 2 つの部品から構成される (Cheng & Lapata, 2016; Nallapati et al., 2017; Cohan et al., 2018)．以後，それぞれの部品について順に説明する．

3.3.1　単語エンコーダ

単語エンコーダは入力文書中の文 x_i に対して，リカレントニューラルネットワークを用いて単語を逐次読み込み，文 x_i を表現するベクトル h_i を得る．具体的には，文 x_i に対し両方向 Long Short Term Memory によるリカレントニューラルネットワークを用いて以下のように h_i を得る．

$$\overrightarrow{e}_{i,j} = LSTM(\overrightarrow{e}_{i,j-1}, emb(w_{i,j})), \tag{17}$$

$$\overleftarrow{e}_{i,j} = LSTM(\overleftarrow{e}_{i,j+1}, emb(w_{i,j})), \tag{18}$$

$$e_{i,j} = [\overrightarrow{e}_{i,j}; \overleftarrow{e}_{i,j}], \tag{19}$$

ここで，$[\overrightarrow{e}_{i,j}; \overleftarrow{e}_{i,j}]$ は前向きおよび後ろ向きの LSTM の隠れ状態ベクトルの結合ベクトルを表現する．また，$emb(\dot{})$ は単語 $w_{i,j}$ の埋め込みベクトルを表す．

加えて，単語エンコーダでは文 x_i の単語間の依存構造も考慮する．Hashimoto ら (Hashimoto & Tsuruoka, 2017) の提案したエンコーダをもとに，文中の任意の単語ペア $(w_{i,k}, w_{i,j})$ に対し以下のように $w_{i,k}$ が $w_{i,j}$ の親になりやすさを計算する．

$$parent(i,j,k) = \frac{\exp(e_{i,j}^T W_{dep} e_{i,k})}{\sum_{j \neq l} \exp(e_{i,j}^T W_{dep} e_{i,l})}, \tag{20}$$

ここで，W_{dep} は重み行列である．文 x_i 中の各単語について，その親を表現するベクトル $z(w_{i,j})$ を以下のように計算する．

$$z(w_{i,j}) = \sum_{k \neq j} parent(i,j,k) \times e_{i,k}. \tag{21}$$

単語エンコーダは最終的に文 x_i に対し，前向きおよび後ろ向きの LSTM の出力を以下のよ

うに結合したベクトル h_i $(1 \leq i \leq N)$ を得る．

$$\overrightarrow{h}_{i,j} = LSTM(\overrightarrow{h}_{i,j-1}, [e_{i,j}; z(w_{i,j})]), \tag{22}$$

$$\overleftarrow{h}_{i,j} = LSTM(\overleftarrow{h}_{i,j+1}, [e_{i,j}; z(w_{i,j})]), \tag{23}$$

$$h_i = [\overrightarrow{h}_{i,M_i}; \overleftarrow{h}_{i,1}]. \tag{24}$$

3.3.2 文エンコーダ

文エンコーダは入力中の各文に対し得られたベクトル h_i を，さらに両方向 LSTM を用いて読み込む．具体的には前向き LSTM$\overrightarrow{H}_i = LSTM(\overrightarrow{H}_{i-1}, h_i)$ および後ろ向き LSTM の $\overleftarrow{H}_i = LSTM(\overleftarrow{H}_{i+1}, h_i)$ を以下のように結合し，文脈を考慮したベクトル H_i を各文 x_i に対し生成する．

$$H_i = [\overrightarrow{H_i}; \overleftarrow{H_i}]. \tag{25}$$

3.3.3 文スコアリング器

文スコアリング器は入力中の文 x_i を要約に含める確率 $p(y_i = 1|\boldsymbol{x}, \theta)$ を出力する．確率値の計算には以下の式を用いる．

$$p(y_i|\mathbf{x}, \theta) = softmax(\mathbf{W}_o H_i). \tag{26}$$

ここで，W_o は重み行列である． この式では，文エンコーダが生成した文脈を考慮する文ベクトル H_i に重み行列をかけ合わせ 2 次元に圧縮する．2 次元ベクトルの各次元がそれぞれ抽出する確信度スコア，および抽出しない確信度スコアに対応する．ソフトマックス関数を用いてこのベクトルの各次元が確率分布となるよう正規化する．

3.4 談話構造を考慮する注意機構

本節では前節までに説明したリカレントニューラルネットワークに基づく抽出型要約モデルに，談話構造を考慮する機構を組み込む方法について説明する．前節で述べた文スコアリング器では，文脈を考慮する文表現 H_i をソフトマックス関数で構成される出力層に入力することにより，文 x_i を抽出する確率値を得た．本研究では，談話構造を考慮するために，出力層に x_i の親を表現するベクトル Ω_i も入力し，談話考慮しながら抽出する確率を計算するモデルを提案する．具体的には，以下のように親を表現するベクトル Ω_i を H_i に結合し出力層へ渡す．

$$p(y_i|\mathbf{x}, \theta) = softmax(\mathbf{W}_o[H_i; \Omega_i]). \tag{27}$$

本節では，まず注意機構の動作の概要について説明し，その後定式化を行う．

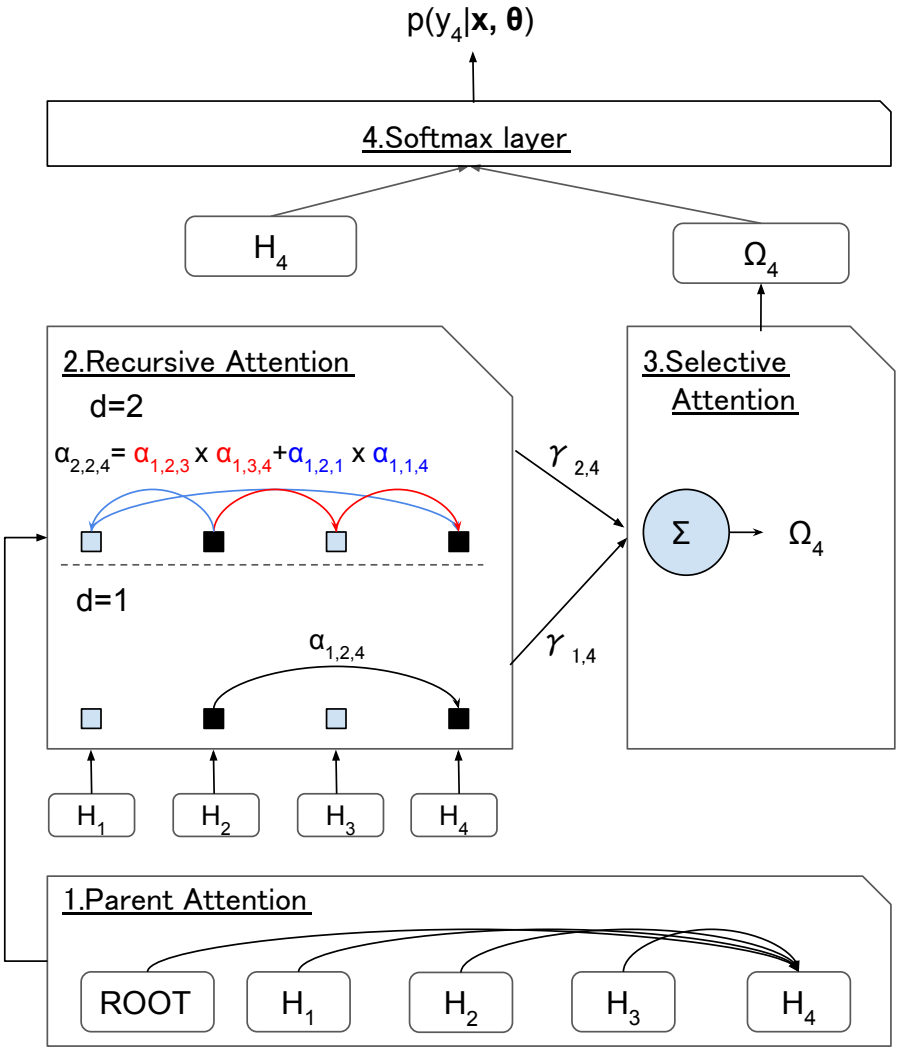


図 3: 談話構造を考慮する注意機構の動作概要.

3.4.1　注意機構の動作概要

Kamigaito ら (Kamigaito et al., 2018) らは，文圧縮課題において入力文中の単語の依存関係を考慮するための注意機構を提案した．本研究では，文 x_i の親を表現するベクトル Ω_i の計算には Kamigaito ら (Kamigaito et al., 2018) と似た手法を採用する．図 3 に Ω_i のための 3 ステップおよび，出力層の概念図を示す．以下に 3 ステップの動作概要を説明する．

ステップ 1: Parent Attention Module このステップでは，入力文書に含まれる文集合から 2 文 (文 k と文 i とする) を選ぶすべての組み合わせについて，文 i が文 k の親になる確率を計算する．ここでは，この確率を $p(k|i, \boldsymbol{H})$ と表現する．図 3 においては，始点

が親を示し，終点が子を示す．$p(k|i, \boldsymbol{H})$ は辺の重みとなり，入力文書はステップ 1 が終了した段階で重み付き全結合グラフとして表現される．

ステップ 2: Recursive Attention Module ステップ 2 では，ステップ 1 で計算した辺の重みを用いて，入力文書中のそれぞれの文 x_i に対して，その d 次親を表現する文ベクトル $\boldsymbol{H}$ に対する荷重合計ベクトル $\gamma_{d,i}$ を生成する．ここで，d 次とは文の組み合わせ文 x_i と文 x_k との距離を表す．

ステップ 2 は，はじめに $d = 1$ の場合の計算から始める．すなわち，x_i と直接的な親子関係を持つ文を表現するベクトルの生成から始める．ステップ 1 において，すでにすべての組み合わせ x_i と x_k について，親になりやすいさを計算している．よって，ステップ 2 においては親になりやすさを表現する確率値を H_k の重み $\alpha_{1,k,i}$ として，$d = 1$ の場合の荷重合計ベクトル $\gamma_{1,i}$ を得る．$\gamma_{1,i}$ は文 x_i と直接的な親子関係を持つベクトルを表現する．

$d = 2$ の場合には，x_i と距離が 2 の親子関係を持つベクトル $\gamma_{2,i}$ を得ることが目標となる．文 x_i 以外のすべての文 x_k に対して，文 x_k が文 x_i の距離 2 の親子関係を持つ確率 $\alpha_{2,k,i}$ を計算し，文ベクトル H を $\alpha_{2,k,i}$ を用いて荷重合計し，$\gamma_{2,i}$ を得る．重み $\alpha_{2,k,i}$ の計算方法について説明する．例えば，図 3 において，文 x_2 から文 x_4 に向かう距離の経路を考えると，赤色の経路と青色の経路が存在する．$\alpha_{2,k,i}$ の計算においては，まず，赤色および青色それぞれの辺に重みを掛け合わせる．すなわち，赤色の経路については $\alpha_{1,2,3} \times \alpha_{1,3,4}$，青色の経路については $\alpha_{1,2,1} \times \alpha_{1,1,4}$ を得る．最後に，かけ合わせた青色および赤色の経路の重みを足し合わせ最終的に $\alpha_{2,2,4}$ を得る．つまり，$\alpha_{2,2,4} = \alpha_{1,2,3} \times \alpha_{1,3,4} + \alpha_{1,2,1} \times \alpha_{1,1,4}$ となる．$\alpha_{2,i,k}$ を用いて文ベクトル H を荷重合計することで，x_i の 2 次の親を表現するベクトル $\gamma_{2,i}$ 得る．

$d > 2$ の場合についても，同様の計算を繰り返すことで再帰的に計算することができる．

ステップ 3: Selective Attention Module ステップ 2 において，次数 d についてそれぞれ親ベクトルを生成している．ステップ 3 においては，これらの複数のベクトルに対し再度重みを計算し荷重合計することで，最終的な親を表現するベクトル Ω_i を得る．

3.4.2 談話構造を考慮する注意機構の定式化

前節において説明したステップ 1，ステップ 2 およびステップ 3 を以下に定式化する．

ステップ 1 の Parent Attention Module は文 x_k が x_i の親になる確率をすべての文の組み合わせについて計算する．

$$p(k|i, \mathbf{H}) = softmax(g(k, i)), \tag{28}$$

$$g(k, i) = v_a^{\mathrm{T}} \tanh(U_a \cdot H_k + W_a H_i), \tag{29}$$

ここで，v_a, U_a および W_a は重み行列である．

ステップ 2 の Recursive Attention Module は x_k が x_i の d 次の親になる確率を以下のように計算する．

$$\alpha_{d,k,i} = \begin{cases} p(k|i, \mathbf{H}) & (d = 1), \quad (30) \\ \sum_{l=0}^{N} \alpha_{d-1,k,l} \times \alpha_{1,l,i} & (d > 1). \quad (31) \end{cases}$$

さらに，談話依存構造木において ROOT は親を持たない制約を以下のように追加する．

$$\alpha_{1,k,i} = \begin{cases} 1 & (k = 0, i = 0), \quad (32) \\ 0 & (k > 0, i = 0), \quad (33) \\ 0 & (k = i, i \neq 0). \quad (34) \end{cases}$$

最初の式では ROOT の親は常に ROOT になるよう制約を与えている．次の式では，ROOT へ入ってくる辺は ROOT のみになるよう制約を加えている．この 2 つの制約により，ROOT は入力文書中の文を親にしない．最後の式は入力文書の各文は自分自身を親にできない制約である．これにより，ROOT がもっとも根に近くなる木構造が学習されるように工夫する．

ステップ 1 において計算された辺の重み $\alpha_{d,k,i}$ を用いて以下のように文ベクトル $\mathbf{H}$ に対する荷重合計ベクトル $\gamma_{d,i}$ を得る:

$$\gamma_{d,i} = \sum_{k=0}^{N} \alpha_{d,k,i} H_k. \tag{35}$$

それぞれの次数に対し，$\gamma_{d,i}$ が得られたら，ステップ 3 ではさらにそれぞれの $\gamma_{d,i}$ に重要度を計算する:

$$\beta_{d,i} = softmax(\mathbf{W}_\beta \gamma_{d,i}), \tag{36}$$

ここで，$\mathbf{W}_\beta$ は重み行列である．

計算された重み $\beta_{d,i}$ を用いて，再度 $\gamma_{d,i}$ を荷重合計して最終的な親ベクトル Ω_i を得る:

$$\Omega_i = \sum_d \beta_{d,i} \cdot \gamma_{d,i}. \tag{37}$$

親を表現するベクトル Ω_i は文 H_i に結合され，出力層によって要約として抽出する確率値 $p(y_i|\mathbf{x},\theta)$ が計算される:

$$p(y_i|\mathbf{x},\theta) = softmax(\mathbf{W}_o[H_i;\Omega_i]). \tag{38}$$

3.4.3 目的関数

学習時には，訓練データに含まれる文書の抽出文と談話構造木をより再現するようにパラメータが更新される．具体的には，以下の式に示す損失関数を最小にするようパラメータを更新する．

$$-\log p(\mathbf{y}|\mathbf{x}) - \lambda \cdot \sum_{k=1}^{N}\sum_{i=1}^{N} E_{k,i} \cdot \log \alpha_{1,k,i}. \tag{39}$$

この式において，$E_{k,i}$ は訓練事例の談話構造木に文 x_i から文 x_k への辺が含まれるなら 1 と取り，含まれないならば 0 になる．したがって，訓練事例に存在する辺に対しより高い重み $\alpha_{1,k,i}$ を与えれば，損失がより小さくなる．第 1 項は既存の要約モデルの学習にも用いられ，抽出文により高い確率値を付与すると損失が小さくなる．すなわち，この 2 つの項で文抽出器と談話構造解析器を同時に学習するようパラメータを更新している．λ はこの 2 つの項の重要度を制御するパラメータである．

3.5 実験

3.5.1 データ

実験には文書要約課題において広く用いられている，DailyMail データセットおよび DUC2002 データセット[1]を用いる．

DailyMail データセットは Herman ら (Hermann, Kocisky, Grefenstette, Espeholt, Kay, Suleyman, & Blunsom, 2015) によって，もともと文書読解課題のために構築された．このデータセットには，Daily Mail Online[2]から収集した新聞記事と，記事の内容を人間が短く 3 文程度のハイライトと呼ばれる記事のまとめが付与されている．Nallapati ら (Nallapati et al., 2016) らはハイライトを要約とみなし，ニューラルネットワークに基づく生成型要約モデルを学習する手法を提案した．抽出文に関するアノテーションはこのデータセットに存在しないため，抽出型要約モデルの学習にそのまま用いることは難しい．そこで，Chen ら (Cheng & Lapata, 2016) はハイライトを参照要約とみなした場合に抽出することで ROUGE 値を最大にする文を抽出文とみなすことで，自動アノテーションした．このデータセットを用いるこ

[1] https://www-nlpir.nist.gov/projects/duc/guidelines/2002.html
[2] http://www.dailymail.co.uk/

とで，抽出型要約モデルの学習も可能になった．Nallapati ら (Nallapati et al., 2017) も抽出型要約モデルを新たに提案し，Chen らの作成したデータセットを用いて実験を行った．現在では，Chen らおよび Nallapati らの提案した抽出型要約モデルをもとに拡張が行われている．本研究でも，彼らのモデルを談話構造を考慮するよう拡張する．

次に DailyMail データセットの前処理について説明する．データセット中のすべての文書に対し，Herman ら (Hernault, Prendinger, Ishizuka, et al., 2010) の談話構造解析器を用いて，談話構造の解析を行った．解析された談話構造はさらに，Hirao ら (Hirao et al., 2013) の手法を用いて，談話依存構造木に変換した．Herman らの談話構造解析器は解析のために段落境界の情報を必要とするが，文書要約課題で共通に用いられる Chen らの配布するデータではこれらの情報が除去されている．よって，本研究では Herman らの提供する元データに対して談話構造解析用いて談話構造の解析を行い，単語の重複の手がかりを用いて，Chen らのデータのデータ分割をできる限り再現した．したがって，196,557 事例の学習データ, 12,147 事例の開発データ，10,396 事例の評価データを獲得した．談話構造解析器が解析できなかった文書などの影響で，訓練データが Chen らの配布するデータよりも少ない．

DUC2002 データセットは 562 文書が含まれており，文書には 2 つの人間による 100 単語以下の要約が付与されている．本研究では要約モデルの学習には DailyMail データセットを用い，DUC2002 データセットは評価にのみ用いる．

3.5.2　評価手法

談話構造の情報を付与した DailyMail データセットを用いて提案モデルの性能を評価する．評価には共通のベンチーマーク指標である ROUGE(Lin, 2004) を用いる．ROUGE にはいくつかの種類が存在するが，既存研究 (Nallapati et al., 2017; Cheng & Lapata, 2016) でも用いられている ROUGE-1, ROUGE-2 および ROUGE-L を用いる．ROUGE-1 は参照要約中のユニグラムを提案手法による要約がどの程度再現できたかを示す，再現率に基づく評価指標になっている．ROUGE-2 はバイグラム，ROUGE-L はもっとも長い系列が参照要約とシステム出力でどの程度一致したかを評価する．ROUGE の計算には Python のパッケージである Pyrouge を用いた[3]．DailyMail データセットのハイライトを参照要約とみなし，提案モデルおよびベースライン手法の ROUGE を比較する．ROUGE は再現率を評価する指標であるため，出力長を長くすると高くなる．本研究では，既存研究の設定にならい 75 バイトおよび 275 バイト (Nallapati et al., 2017; Cheng & Lapata, 2016) および正解要約長を出力長制

[3]ROUGE の計算時のオプションには “-a -c 95 -m -n 2 -b 75” and “-a -c 95 -m -n 2 -b 275” を用いた．

約として用いる．最後の出力長制約については，理想的な出力長制約のもとでどの程度システムが必要な情報を埋め込むことができるかを評価する．

DUC2002 での ROUGE については，学習に用いたデータとは異なるドメインでの評価として報告する．すなわち，提案モデルや比較手法は DailyMail データセットで学習され，DUC2002 データセットで評価する．DUC2002 の公式ガイドラインに沿い，ROUGE-1, ROUGE-2 および ROUGE-L を報告する．

3.5.3 モデル学習

ニューラルネットワークの最適化には Adam(Kingma & Ba, 2015) を採用し，学習率は 0.001 に設定した．また，提案モデルの制御パラメータ λ は 0.1 に設定した．Nallapati ら (Nallapati et al., 2017) の実験設定に従い，語彙サイズを 150,000 に制限し，辞書に含まれない単語は `UNK` という特別なトークンで置換した．バッチサイズは 8 に設定した．LSTM や単語埋め込み層の次元はともに 100 に設定し，単語埋め込み層は word2vec-slim[4]を用いて初期化した．なお，単語埋め込み層のパラメータは他のパラメータと同様に学習中にも更新する．学習を効率化するために，Nallapati ら (Nallapati et al., 2017) にならい 50 文以上から構成される文書を訓練データから除いた．

3.5.4 比較手法

実験では，以下に示す手法を比較する．

`DIS w/ PAR` 制御パラメータ λ の値を 0.01 に設定した提案モデルである．この設定においては，訓練事例における文抽出アノテーションに加え，談話構造解析器によって付与された談話構造木も再現するよう，パラメータが学習される．評価時には，出力層が付与した確率値が大きな順に文を並べ替え，出力長制約を満たすまで先頭から順に文を抽出する．同様の文選択手法は Nallapati ら (Nallapati et al., 2017; Cheng & Lapata, 2016) の研究でも用いられている．

`DIS w/o PAR` 制御パラメータ λ を 0 に設定した提案モデルである．この設定では，損失関数において談話構造解析器の出力した談話構造を再現する項が除去される．よって，すべてのパラメータは文抽出器の性能を最大化するように更新される．この手法においても同様に，出力層の出力する確率値が大きな順に文を並べ替え，出力長制約を満たすまで文を選択する．

[4]https://github.com/eyaler/word2vec-slim

`Lead-3` 先頭 3 文を抽出するベースライン手法である．出力長制約を満たさない場合は，制約を満たすように長い部分を除去する．このベースラインは既存の要約研究において，強いベースライン手法として知られている．

`SummaRuNNer` Nallapati ら (Nallapati et al., 2017) によるリカレントネットワークに基づく抽出型要約手法である．エンコーダは単語エンコーダと文エンコーダから構成されており，本研究での構成と同様である．我々のモデルとの違いは，出力層に入力する情報である．我々のモデルでは，対象文のベクトルに加えて親を表現したベクトルを出力層の入力とする．Nallapati らの手法では，対象文の重要度，情報の真新しさ，文書中での位置などを表現したベクトルを出力層に入力する．この手法はリカレントニューラルネットワークで構成される強い手法として知られている．

`NeuralSum` Chen ら (Cheng & Lapata, 2016) による畳み込みニューラルネットワークを用いたモデルである．このモデルでは，エンコーダとしてリカレントニューラルネットワークではなくて，畳み込みニューラルネットワークを用いる点で本研究と異なる．また，本研究とは違い LSTM を用いたデコーダが採用されている．本研究での提案モデルや `SummaRuNNer` は LSTM デコーダではなく，ソフトマックス関数に基づく出力層にエンコーダが生成したベクトルをそのまま入力する．

`Refresh` Narayan ら (Narayan, Cohen, & Lapata, 2018) の提案した強化学習とエンコーダデコーダを組み合わせる手法である．この手法では，エンコーダデコーダに基づく抽出型手法に ROUGE 値を最大化する報酬を同時に最適化する報酬を加えることで高い ROUGE 値を達成している．

上述の手法に加えて，DUC2002 データセットでの評価においてはいくつかの手法との比較を追加で行う．`LREG` (Cheng & Lapata, 2016) はロジスティック回帰に基づく手法で，いくつかの素性を用いて文をベクトル化して，抽出するか否かの確率値を回帰問題として定式化する．`ILP` (Woodsend & Lapata, 2010) は制約のもとで整数計画法を解くモデルである．この手法は抽出単位が文ではなく節となり，出力長や文法性などの制約を満たす要約を出力する．`TGRAPH` (Parveen et al., 2015) と `URANK` (Wan, 2010) はグラフに基づく要約手法であり，DUC2002 データセットでは良い性能を示すことが知られている．

3.6　結果および考察

	75			275			正解要約長		
	R-1	R-2	R-L	R-1	R-2	R-L	R-1	R-2	R-L
DIS w/ PAR, d={1}	22.9	9.6	12.1	41.9	17.4	**35.2**	$^{+}$41.1	$^{+}$**16.9**	$^{+}$36.7
DIS w/ PAR, d={1,2}	$^{+}$**25.0**	$^{+}$11.1	$^{+}$13.4	41.7	17.4	35.0	41.0	16.7	$^{+}$36.7
DIS w/ PAR, d={1,2,3}$^{\#}$	$^{+}$24.6	$^{+}$**11.2**	$^{+}$**13.5**	41.8	17.2	**35.2**	$^{+}$41.1	$^{+}$16.8	$^{+}$36.8
DIS w/ PAR, d={1,2,3,4}	$^{+}$24.1	10.8	$^{+}$13.2	41.1	**17.5**	35.1	$^{+}$**41.2**	$^{+}$**16.9**	$^{+}$**37.0**
DIS fixed, d={1}	23.4	10.3	12.7	39.9	16.1	33.5	39.4	15.7	35.4
DIS fixed, d={1,2}	23.5	10.4	12.8	40.3	15.9	33.6	39.7	16.1	35.8
DIS fixed, d={1,2,3}	22.9	9.8	12.3	40.3	16.4	33.8	39.6	15.9	35.6
DIS fixed, d={1,2,3,4}	22.6	9.2	11.7	39.8	15.6	33.4	39.3	16.1	35.9
DIS w/o PAR, d={1}	21.2	8.1	11.0	40.1	15.8	33.7	39.6	15.5	35.5
DIS w/o PAR, d={1,2}	21.1	7.5	10.6	40.0	15.8	33.0	39.6	15.6	35.5
DIS w/o PAR, d={1,2,3}	20.9	7.9	10.9	40.5	16.1	34.1	40.0	15.8	35.8
DIS w/o PAR, d={1,2,3,4}	21.1	8.0	10.9	40.2	15.7	33.6	39.6	15.5	35.6
Lead-3	23.0	9.4	11.8	41.9	17.0	32.5	40.4	16.3	36.1
no-attn	20.1	7.1	10.4	39.6	15.4	33.3	39.3	15.3	35.2
SummaRuNNer (re-run)	23.2	9.6	11.0	**42.0**	17.2	32.5	37.6	14.8	33.7
Refresh (re-run)	23.1	10.9	12.6	37.9	16.5	31.4	36.6	15.8	34.1
NeuralSum (re-run)	22.4	9.1	11.8	40.8	16.3	34.8	40.3	15.9	36.1

表 1: DailyMail データセットによる ROUGE 値による自動評価．DailyMail データセットで学習し，DailyMail データセットで評価．太字は最も良い性能を達成したモデルを示す．+記号はもっとも性能が近いベースライン手法と統計的有意な差を得たモデルである．#記号は開発セットでの ROUGE-2 が最大となったモデルである．

3.6.1 DailyMail データセットでの結果

表 1 に比較手法の ROUGE 値を示す．表の各カラムはそれぞれ 75 バイト，275 バイトおよび正解出力長での ROUGE-1，ROUGE-2 および ROUGE-L となっている．この表において，既存論文の報告値にはアスタリスクがついている．既存論文ではデータセット中の固有名詞などを@entity などの特別なトークンで置換したものを用いており，本研究と使用しているデータセットが厳密には異なる．本研究では，既存手法の再実装を用い，我々のデータと同じデータで学習したモデルと比較している．

はじめに，出力長制約が 75 バイトの場合の結果について述べる．この設定では提案モデルである DIS w/ PAR (d={1,2} および DIS w/ PAR (d={1,2,3} がすべての ROUGE 値において，談話構造を考慮しない比較手法よりも良い性能を示した．これらの値は談話構造を考慮しないベースライン手法と統計的に有意な差を得ている．ROUGE-2 は人間による評価との相関がより良いことが知られており，75 バイトの出力長制約において提案モデルはもっとも良い

表 2: DUC2002 データセットでの ROUGE 値

	R-1	R-2	R-L
LEAD-3	43.6	21.0	40.2
LREG	43.8	20.7	40.3
ILP	45.4	21.3	42.8
TGRAPH	48.1	**24.3**	-
URANK	**48.5**	21.5	-
SummaRuNNer	46.6	23.1	43.0
NeuralSum	47.4	23.0	43.5
DIS w/ PAR	44.7	21.8	38.2
DIS w/o PAR	46.1	23.3	**43.6**

要約器となっている．DIS w/o PAR は SummaRuNNer と有意差は見られず，Lead-3 よりも良い性能を示した．DIS w/o PAR と DIS w/ PAR を比較すると，DIS w/ PAR のほうが良い性能を示した．このことから，λ の値を 0 よりも大きくし，談話構造解析器のアノーテションした木構造を再現するようにパラメータを学習することで，性能をあげられることがわかった．

次に 275 バイトの出力長設定での結果について述べる．この設定においては，SummaRuNNer が ROUGE-1 においてもっとも良い性能を示した．ROUGE-2 および ROUGE-L については，提案手法である DIS w/ PAR $d = 1, 2, 3, 4$ と DIS w/ PAR $d = 1, 2, 3$ が SummaRuNNer よりも良い性能を示した．しなしながら，これらの手法はどれも近い性能を示しており統計的な有意差は確認できない．要約長制約が長い設定では，談話構造を考慮する機構の効果は限定的である．

出力長制約を正解要約に合わせた場合の ROUGE 値について考察する．この制約において，ROUGE 値は 275 バイト制約のときと似た傾向を示した．すなわち，すべての比較手法が似たような ROUGE 値となった．この結果から，談話構造の情報を明示的に組み込むと，とくに短い出力長制約の設定において有利に働くことがわかる．

3.6.2　DUC2002 データセットでの結果

次に DUC2002 データセットでの性能評価について考察する．表 2 に各比較手法の ROUGE-1，ROUGE-2 および ROUGE-L を示す．ニューラルネットワークに基づく手法はどれも近い ROUGE 値を示した．DUC2002 データセットは 100 単語程度で参照要約が作られていることから，75 バイトよりは長い．よって，DailyMail データセットでの結果と同様に長い出力長

制約であり，差が見られなかった．一方で，TGRAPH や URANK のようなグラフに基づく手法がニューラルネットワークに基づく手法よりも良い性能を示している．Nallapati ら (Nallapati et al., 2017) が考察しているように，ニューラルネットワークに基づく手法は学習に用いたドメインのデータに強く適応し，ドメイン以外のデータに対しては大きく性能が劣化するものと思われる．また，もう一つ考えられる理由としてはデータセット中の抽出文へのアノテーション方法が挙げられる．Chen ら (Cheng & Lapata, 2016) のデータ作成手法においては，DailyMail データセットのハイライトを参照要約とみなし，ROUGE-F を最大にするよう抽出文をアノテーションする．DUC2002 データセットの参照要約は DailyMail データセットのハイライトよりも長い．よって，出力長制約が異なる学習データから学習したことにより，性能が劣化した可能性がある．

3.6.3 人手評価

本研究では ROUGE 値による自動評価に加え，人間の評価者による 5 段階採点による人手評価も行った．DailyMail データセットからランダムに抽出した 100 事例を用い，評価者に情報量および一貫性の観点から 1 点 (もっとも悪い) から 5 点 (もっとも良い) を付与させた．1 事例につき 4 人の評価者が採点を行い，平均点を算出した．全事例での各手法の平均点を表 3 に示す．提案手法である DIS w/ PAR は，一貫性の観点において Lead-3 の次に良い点数を得た．Lead-3 は連続した 3 文を抽出する制約が付いており，本質的に一貫性に観点において非常に強い．DIS w/ PAR はその他の比較手法 (SummaRuNNer および DIS w/o PAR) よりも一貫性において統計的有意に良い点数を得た．ROUGE 値による自動評価においては，統計的な有意差を得られなかった要約長制約が 275 バイトおよび正解要約長の設定においても，人手評価の場合は一貫性において比較手法よりも統計的に有意に良い評価を得た．比較手法の中でも SummaRuNNer はとくに一貫性において低い点を得ており，ROUGE 値は高いが要約としては一貫性の観点では品質が低くなっている．

3.6.4 定性的分析

表 4 に入力文書と提案モデル DIS w/ PAR によって抽出された文を例示する．表において，提案モデルによって抽出された文は太字で示す．また，図 4 に学習された談話依存構造木を示す．この図において，ノードは文を表現し，要約に含められ文は赤で示す．図に示す木は，木になる制約を満たしながら辺の重みの総和が大きくなるよう含める辺を貪欲法により探索したものである．辺の重みは Parent Attention が計算した親子関係になりやすいさを表す確率値である．太い辺は提案手法によって同定された構造と談話構造解析器によるアノテー

	75		275		Ref	
	一貫性	情報量	一貫性	情報量	一貫性	情報量
DIS w/ PAR	*3.86	3.57	***4.11**	***3.97**	***3.98**	3.78
SummaRuNNer	3.61	3.57	2.98	3.77	2.81	3.16
no-attn	3.73	3.52	3.92	3.86	3.80	3.70
LEAD-3	**3.98**	**3.69**	4.06	**3.97**	3.94	**3.80**

表 3: ランダムに抽出した 100 事例を用いた人手評価

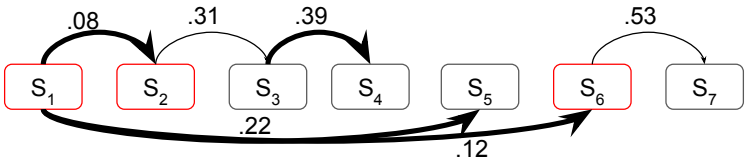


図 4: 学習された木構造の例

ションが同一のものである．

表 4 の太字の文と参照要約を比較すると，正解に近い文が抽出できている．図 4 の辺に着目すると，6 つの辺のうち 4 つの辺が談話構造解析器の出力した辺と一致している．談話構造解析器の情報は学習時にのみ用い，評価時には談話構造のアノテーションを用いないにもかかわらず，提案手法はこの例において談話構造解析器の出力に近い木構造を出力した．また，抽出された文に着目すると，ROOT に近いノードから選択されている．よって，重要な文ほど木構造の ROOT に近い部分に集まっており，抽出文を決定する際の手がかりとして有効に機能しているようにみえる．また，$S1$ から $S6$ へのエッジなど長距離の依存関係を正しく捉えられている，従来の LSTM のみを用いるエンコーダではこのような長距離の依存関係を捉えるのが難しかったが，提案モデルではこのような長期依存関係も含めて出力層に入力できるため，性能が向上したものと思われる．

3.7　談話構造を考慮する要約モデルのまとめ

本研究では入力文書の談話構造を考慮する要約モデルを提案した．リカレントニューラルネットワークに基づく抽出型要約器に，入力文書の木構造を捉える注意機構を新たに導入した．実験より，提案モデルが ROUGE 値を用いた自動評価において現在良い性能を示している既存モデルよりもより良いか，同等の性能を示した．人手評価においてはとくに一貫性の

表 4: 出力要約の例.

Document:
S1: Bayern Munich have an interest in Chelsea defender Branislav Ivanovic but are unlikely to make a move until January.
S2: The 31-year-old Serbia captain has yet to open talks over a new contract at Chelsea and his current deal runs out in 2016.
S3: Chelsea defender Branislav Ivanovic could be targeted by Bayern Munich in the January transfer window.
S4: Bayern like Ivanovic but don't expect Chelsea to sell yet they know he will be free to talk to foreign clubs from January.
S5: Paris Saint-germain will make a £ 7million offer for Chelsea goalkeeper . Petr Cech this summer
S6: The 32-year-old is poised to leave Stamford Bridge and wants to play for a champions league.
S7: Contender PSG are set to make a £ 7million bid for Ivanovic's Chelse a team-mate Petr Cech in the summer.
Gold Summary:
Branislav Ivanovic's contract at Chelsea expires at the end of next season. The 31-year-old has yet to open talks over a new deal at Stanford bridge. Petrcech is poised to leave Chelsea at the end of the season.

観点から既存手法よりも良い性能を得た．これより，談話構造を考慮する注意機構の効果を出来た．

第4章　質問を対象とする要約課題の提案とニューラル要約モデルの適用

本章では，要約研究の二つ目の方向である要約対象テキストの多様化に対し行った研究について説明する．

4.1　研究概要

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! Answers[1]から抜粋した質問の例を表 5 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 5 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．

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 & Carenini, 2014; Oya et al., 2014)

[1]https://answers.yahoo.com/

表 5: 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.

質問を対象とする要約研究としては Tamura ら (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．

4.2 質問要約課題の特徴

本研究で提案する "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．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．単一文質問への要約課題に限定することで，例えば，

[2]https://duc.nist.gov

出力した要約をコミュニティ質問応答サイトのタイトルとしてそのまま使うことができるなど実応用の観点から利点がある．

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．

1. 質問文と叙述文が混在する (**個別文の性質の違い**)
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，文書中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，個別文の性質や談話構造に関し従来の要約課題とは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 5 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? (塩素によって髪の染料が落ちますか？)” といった疑問文だけでなく，“Affect of the chlorine on my hair eyeing. (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．

表 6: 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

表 7: 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

4.3 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 1.0[3]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-要約対を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．
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．

4.3.1 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

[3]https://webscope.sandbox.yahoo.com/

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 6 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．

4.3.2　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．

4.3.3　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 7 に示す．また，代表的な質問テキスト-タイトル対の例を表 8 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 8 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．

4.4 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．

4.4.1 データセット

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 9 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．データセット全体の 4,483,031 投稿うち，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 Bordaux 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．

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 Turk[4]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 10 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 “中国に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 およびタイトルは食べ物にはフォーカスしない要約となっている．

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，タイトルどちらも本文中の表現をしばしば言い換える．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

[4]https://www.mturk.com/

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．

4.4.2　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．

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 8 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 8 の例 4 において，“Why?”，末尾疑問文は “What am I dong 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．

機械学習に基づく抽出型手法

機械学習に基づく抽出型手法としては，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．機械学習に基づく抽出型手法についても，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，まず入力の各文に対し ROUGE-2 F 値の予測値を出力する Support Vector Regression (SVR) (Basak, Pal, & Patranabis, 2007) を学習する．学習済みの回帰モデルを用いて，入力の各文に対し 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 & Vandewalle, 1999) を学習する．学習した SVM を用い，入力の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．

4.4.3 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 ら (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 ら (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 $\boldsymbol{x} = (x_1, \ldots, x_n)$ 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 $\boldsymbol{h}_\tau$ に，Recurrent Neural Network (RNN) を用いて逐次変換する:

$$\boldsymbol{h}_\tau = f(x_\tau, \boldsymbol{h}_{\tau-1}). \tag{40}$$

f には Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) や Gated Recurrent Unit (GRU) (Cho, van Merrienboer, Gulcehre, Bahdanau, Bougares, Schwenk, & 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 $\boldsymbol{s}_t$ 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 $\boldsymbol{h}_n$ を用いることにする:

$$\boldsymbol{s}_t = f(y_{t-1}, s_{t-1}), \tag{41}$$

$$p(y_t|y_{<t}, \boldsymbol{x}) = softmax(g(\boldsymbol{s}_t)). \tag{42}$$

$\boldsymbol{x}$ が与えられた上での，出力要約 $\boldsymbol{y}$ 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:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{m} p(y_t|y_{<t}, \boldsymbol{x}). \tag{43}$$

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:

$$\log p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{t=1}^{m} \log p(y_t|y_{<t}, \boldsymbol{x}). \tag{44}$$

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，入力の文書を逐次エンコードし，最終状態 $\boldsymbol{h}_n$ を文脈ベクトル $\boldsymbol{c}$ 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 $\boldsymbol{c_t}$ を考える:

$$\boldsymbol{c_t} = \sum_{\tau=1}^{n} \alpha_{t\tau} \boldsymbol{h}_\tau. \tag{45}$$

$\alpha_{t\tau}$ 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:

$$\alpha_{t\tau} = \frac{exp(\boldsymbol{s}_t \cdot \boldsymbol{h}_\tau)}{\sum_{h'} exp(\boldsymbol{s_t} \cdot \boldsymbol{h'})}. \tag{46}$$

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

$$\tilde{\boldsymbol{h}} = tanh(\boldsymbol{W_c}[\boldsymbol{c_t}; \boldsymbol{h_t}]), \tag{47}$$

$$p(y_t|y_{<t}, \boldsymbol{x}) = softmax(\boldsymbol{W_s}\tilde{\boldsymbol{h}}_{\boldsymbol{t}}). \tag{48}$$

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:

$$p(y_t|y_{<t}, \boldsymbol{x}) = score_{gen}(y_t|\boldsymbol{s}_t, y_{t-1}, \boldsymbol{c}_t, \boldsymbol{x}) + score_{copy}(y_t|\boldsymbol{s}_t, y_{t-1}, \boldsymbol{c}_t, \boldsymbol{x}). \tag{49}$$

$score_{copy}$ 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 ($y_t \in \boldsymbol{x}$) 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 ($y_t \notin \boldsymbol{x}$) の場合には 0 を出力する:

$$score_{copy}(y_t|.) = \begin{cases} \frac{1}{Z} \sum_{j:x_j=y_t} \exp(\phi_c(x_j)) & (y_t \in \boldsymbol{x}) \\ 0 & (y_t \notin \boldsymbol{x}). \end{cases} \tag{50}$$

$score_{gen}$ 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 $\boldsymbol{x}$ に含まれる場合には 0 を出力する:

$$score_{gen}(y_t|.) = \begin{cases} \frac{1}{Z}\exp(\phi_g(y_t)) & (y_t \in \boldsymbol{V}) \\ \frac{1}{Z}\exp(\phi_g(UNK)) & (y_t \notin \boldsymbol{x} \quad \wedge \quad y_t \notin \boldsymbol{V}) \\ 0 & (y_t \in \boldsymbol{x} \quad \wedge \quad y_t \notin \boldsymbol{V}). \end{cases} \tag{51}$$

$score_{copy}$ は $y_t \in \boldsymbol{x}$ のときにスコアを出力し $y_t \notin \boldsymbol{x}$ 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 $Z = \sum_{v \in V \cup \{UNK\}} \exp(\phi_g(v)) + \sum_{x \in \boldsymbol{x}} \exp(\phi_c(x))$ 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．

4.5　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．

4.5.1　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, & 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．

4.5.2 ROUGE 値による自動評価

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 11 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かだか 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 ら (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．

4.5.3　人手評価

本研究では ROUGE 値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 Crowdflower[5]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，英語として正しく，入力の質問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 12，13 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

[5]http://crowdflower.com

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．

4.5.4 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 14 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．生成型の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., 2015)，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．

4.6 質問要約のまとめ

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．

表 8: 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 appartment 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 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?

表 9: Yahoo! Answers dataset 全体に含まれる質問テキストの長さ (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．)

質問テキスト中の文数	0	1	2	3	4	5	6	7 以上
文書数 (×1000)	2,291	774	434	297	202	136	92	354

表 10: 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:

1. Staying health during adventure in China.
2. when traveling to china what precautions should be taken before going and eating there?
3. Any precautions when traveling in China?
4. What are some food tips for staying healthy on my trip to China?
5. What are the risks of illness while traveling to China?

表 11: ROUGE-2 による自動評価

	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**

表 12: 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	69	79	82	-

表 13: 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	89	107	103	-

表 14: 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?

第5章　結論と今後の課題

本章では本研究についてまとめ，今後の課題について議論する．

5.1　結論

本研究では要約研究の2つの方向について，ニューラルネットワークに基づく手法という観点から貢献した．

1つ目は要約手法そのものの高度化に対する貢献である．単一文書要約課題にニューラルネットワークに基づく手法が適用されて以来，ニューラルネットワークに基づく抽出型要約モデルにおいては入力文書は文の系列として表現されてきた．一方で，ニューラルネットワークが適用される以前の例えば線形計画法に基づく要約器などにおいては文間の意味的な関係を木構造で表現する談話構造に基づく情報が有効に機能することが示されてきた．本研究では，ニューラルネットワークに基づく抽出型要約器に談話構造を考慮する注意機構を導入することで性能が向上することを示した．具体的には，出力層に対し対象文だけではなくその親を考慮するベクトルも合わせて入力することで木構造を考慮しながら，対象文を要約に含める確率を効果的に計算できるよう工夫した．DailyMail データセットによる評価において，出力長制約が75バイトの設定において既存手法を統計的有意に上回るROUGE値を示した．DUC2002データセットにおいては，提案手法やその他のニューラルネットワークに基づく手法の性能は既存のグラフに基づく要約手法よりも低く，学習に用いたドメイン外のデータに対しての性能劣化を防ぐ手法が今後必要だと考えられる．

2つ目は要約対象テキストの多様化に対する貢献である．要約手法そのものの高度はについては，新聞記事や科学論文を基にした共通の評価用データセットを用いて評価されてきた．一方で，Eメールや会話テキストなど新聞記事や科学論文とは異なる特徴を持つテキストを対象とする要約課題が登場している．本研究では，コミュニティ質問応答サイトなどの登場により爆発的にその量が増えている質問テキストに着目して，質問を対象とする新たな要約課題を提案した．質問テキストについては，ユーザに提示するための要約研究が行われておらず，そもそもどのような要約手法が有効に機能するかが明らかではない．そこで，本研究ではどのような要約手法が効果的に機能するかデータの分析を行い，さらに実際にいくつかの要約手法を適用しその性能を比較した．コミュニティ質問応答サイトの質問投稿から自動

作成したデータセットを構築および分析し抽出型手法では不十分で生成型手法を必要とする質問テキストの存在を確認した．また，規則に基づく抽出型手法，機械学習に基づく抽出型手法およびニューラルネットワークに基づく生成型手法を比較し，ニューラルネットワークに基づく生成型手法が質問要約課題において良い性能を示すことを確認した．

5.2　今後の課題

要約課題の今後の展望について述べる．

1 つ目の研究で示したように，ニューラルネットワークに基づく要約モデルは学習に用いたデータ以外のドメインに適用すると，性能が劣化する．例えば，DailyMail データセットで学習した要約モデルを DUC2002 データセットに適用すると，既存のグラフに基づく要約モデルよりも性能が低くなる．今後の方向として，学習に用いたデータ以外のデータにも頑健に動作する要約器を構築する必要があると考える．具体的には DailyMail データセットなどの大規模なデータで学習したニューラルネットワークを，DUC2002 などの小規模なデータセットで Fine-tune するなどの方法が考えられる．

これまで述べてきたように出力要約はしばしば文法性，一貫性などの観点から評価される．現状の生成型文書要約課題では，出力の文法性は良いものの文間のつながりの良さを表す一貫性についてはさらなる向上の余地があると考える．本研究で提案したのは，生成型手法のデコード時に一貫性が良くなるような注意機構であった．最近では強化学習を用いて生成されるテキストを制御する取り組みも登場してきており，一貫性をデコード時に評価するような報酬を組み込んだ強化学習をに基づく生成型要約モデルが考えられる．

生成型要約は辞書中の単語を選択する問題を繰り返し解くことで要約を生成する．そのため，文全体としてそもそも意味の通らない要約や，差別的な表現など問題のある出力が生成される可能性がある．実応用上の観点からはこのような出力を避ける機構を導入する必要がある．

2 つ目の貢献として取り組んだ質問要約課題においては，既存手法を質問テキストに適用し比較した．比較に用いたニューラルネットワークに基づく手法は，質問を対象として設計されたものではない．そこで，ニューラルネットワークに基づく要約器に質問特有の素性を組み込むことで，質問に特化したニューラル質問要約モデルが構築できるものと考えられる．例えば，入力中の疑問文にはより重要な情報が含まれていると仮定し，疑問文により着目しながらデコードする生成型要約器などが考えられる．

これまで多くの要約研究は入力としてテキストのみを想定してきた．ウェブ上に存在する

文書には画像や動画などマルチモーダルな情報も含まれていることがある．このようなテキスト以外の情報も考慮しながら，テキスト中の重要箇所を同定する “マルチモーダル要約” も発展の余地があると考える．

参考文献

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of International Conference on Learning Representations 2015 (ICLR2015)*.

Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews*, **11** (10), 203–224.

Cheng, J. & Lapata, M. (2016). Neural Summarization by Extracting Sentences and Words. In *Proceedings of 54th Annual Meeting of the Association for Computational Linguistics (ACL2016)*, pp. 484–494 Berlin, Germany.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP2014)*, pp. 1724–1734.

Cohan, A., Dernoncourt, F., Kim, D. S., Bui, T., Kim, S., Chang, W., & Goharian, N. (2018). A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents. In *Proceedings of 2018 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL2018)*, pp. 615–621.

Dorr, B., Zajic, D., & Schwartz, R. (2003). Hedge trimmer: A Parse-and-trim Approach to Headline Generation. In *Proceedings of NAACL 2003 Workshop on Text Summarization*, pp. 1–8.

Duboue, P. A. (2012). Extractive email thread summarization: Can we do better than He said She said?. In *Proceedings of 5th International Conference on Natural Language Generation (INLG2012)*, pp. 85–89.

Erkan, G. & Radev, D. R. (2004). Lexrank: Graph-based Lexical Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research*, **22**, 457–479.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification. In *Journal of Machine Learning Research 9(2008)*, pp. 1871–1874.

Filippova, K., Alfonseca, E., Colmenares, C. A., Kaiser, L., & Vinyals, O. (2015). Sentence Compression by Deletion with LSTMs. In *Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pp. 360–368.

Gu, J., Lu, Z., Li, H., & Li, V. O. (2016). Incorporating Copying Mechanism in Sequence-to-Sequence Learning. In *Proceedings of 54th Annual Meeting of the Association for Computational Linguistics (ACL2016)*.

Hashimoto, K. & Tsuruoka, Y. (2017). Neural Machine Translation with Source-Side Latent Graph Parsing. In *Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP2017)*, pp. 125–135.

Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., & Blunsom, P. (2015). Teaching machines to read and comprehend. In *Proceedings of Twenty-ninth Conference on Neural Information Processing Systems (NIPS2015)*, pp. 1693–1701.

Hernault, H., Prendinger, H., Ishizuka, M., et. al (2010). HILDA: A discourse parser using support vector machine classification. *Dialogue & Discourse*, **1** (3).

Hirao, T., Isozaki, H., Maeda, E., & Matsumoto, Y. (2002). Extracting Important Sentences with Support Vector Machines. In *Proceedings of 19th International Conference on Computational Linguistics (COLING2002)*, pp. 1–7.

Hirao, T., Nishino, M., & Nagata, M. (2017). Oracle summaries of compressive summarization. In *Proceedings of 55th Annual Meeting of the Association for Computational Linguistics (ACL2017)*, pp. 275–280.

Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., & Nagata, M. (2013). Single-document Summarization as a Tree Knapsack Problem. In *Proceedings of 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP2013)*, pp. 1515–1520.

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, **9** (8), 1735–1780.

Kamigaito, H., Hayashi, K., Hirao, T., & Nagata, M. (2018). Higher-Order Syntactic Attention Network for Longer Sentence Compression. In *Proceedings of 2018 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL2018)*, pp. 1716–1726.

Kikuchi, Y., Hirao, T., Takamura, H., Okumura, M., & Nagata, M. (2014). Single document summarization based on nested tree structure. In *Proceedings of 2014 Annual Conference of the North American Chapter of the Association for Computational Linguistics (ACL2014)*, Vol. 2, pp. 315–320.

Kingma, D. P. & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of International Conference on Learning Representations 2015 (ICLR2015)*.

Kobayashi, H., Noguchi, M., & Yatsuka, T. (2015). Summarization based on Embedding Distributions. In *Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pp. 1984–1989.

Li, C., Qian, X., & Liu, Y. (2013). Using Supervised Bigram-based ILP for Extractive Summarization. In *Proceedings of 51st Annual Meeting of the Association for Computational Linguistics (ACL2013)*, pp. 1004–1013.

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of ACL2004 Workshop on Text Summarization Branches Out*, pp. 74–81.

Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM J. Res. Dev.*, **2** (2), 159–165.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pp. 1412–1421.

Mann, W. C. & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, **8** (3), 243–281.

McKeown, K. & Radev, D. R. (1995). Generating Summaries of Multiple News Articles. In *Proceedings of 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR95)*, pp. 74–82.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems*, pp. 3111–3119.

Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents.. In *Proceedings of 21st AAAI Conference on Artificial Intelligence (AAAI2017)*, pp. 3075–3081.

Nallapati, R., Zhou, B., dos Santos, C., glar Gulehrea, & Xiang, B. (2016). Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In *Proceedings of 2016 SIGNLL Conference on Computational Natural Language Learning (CoNLL2016)*, pp. 280–290.

Narayan, S., Cohen, S. B., & Lapata, M. (2018). Ranking Sentences for Extractive Summarization with Reinforcement Learning. In *Proceedings of 2018 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL2018)*, pp. 1747–1759.

Oya, T. & Carenini, G. (2014). Extractive Summarization and Dialogue Act Modeling on Email Threads: An Integrated Probalistic Approach. In *Proceedings of 15th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL2014)*, pp. 133–140.

Oya, T., Mehdad, Y., Carenini, G., & Ng, R. (2014). A Template-based Abstractive Meeting Summarization: Leveraging Summary and Source Text Relationships. In *Proceedings of 7th International Conference on Natural Language Generation (INLG2014)*, pp. 45–53.

Parveen, D., Ramsl, H.-M., & Strube, M. (2015). Topical coherence for graph-based extractive summarization. In *Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pp. 1949–1954.

Peyrard, M. & Eckale-Kohler, J. (2016). Optimizing an Approximation of ROUGE- a Problem-Reduction Approach to Extractive Multi-Document Summarization. In *Proceedings of 54th Annual Meeting of the Association for Computational Linguistics (ACL2016)*, pp. 1825–1836.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., & Webber, B. L. (2008). The Penn Discourse TreeBank 2.0.. In *the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Rush, A. M., Chopra, S., & Weston, J. (2015). A Neural Attention Model for Sentence Summarization. In *Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pp. 379–389.

See, A., Liu, P. J., & Manning, C. D. (2017). Get To The Point: Summarization with Pointer-Generator Networks. In *Proceedings of 55th Annual Meeting of the Association for Computational Linguistics (ACL2017)*, Vol. 1, pp. 1073–1083.

Shen, D., Sun, J.-T., Li, H., Yang, Q., & Chen, Z. (2007). Document Summarization using Conditional Random Fields. In *Proceedings of IJCAI2007*, pp. 2862–2867.

Suykens, J. A. & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, **9** (3), 293–300.

Tamura, A., Takamura, H., & Okumura, M. (2005). Classification of Multiple-Sentence Questions. In *Proceedings of 2nd International Joint Conference on Natural Language Processing (IJCNLP2005)*, pp. 426–437.

Wan, X. (2010). Towards a unified approach to simultaneous single-document and multi-document summarizations. In *Proceedings of 27th International Conference on Computational Linguistics (COLING2010)*, pp. 1137–1145.

Woodsend, K. & Lapata, M. (2010). Automatic generation of story highlights. In *Proceedings of 48th Annual Meeting of the Association for Computational Linguistics (ACL2010)*, pp. 565–574.

研究業績

論文誌

- 石垣 達也, 高村大也, 奥村学, 複数文質問を対象とした抽出型および生成型要約，自然言語処理, **26** (1), pp. 37–58, 2019

国際会議論文

- Tatsuya Ishigaki, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura, Discourse-aware Hierarchical Attention Network for Extractive Single-Document Summarization, In *Proceedings of Recent Advances In Natural Language Processing (RANLP2019)*, 2019
- Tatsuya Aoki, Akira Miyazawa, Kasumi Aoki, Keiichi Goshima, Tatsuya Ishigaki, Ichiro Kobayashi, Hiroya Takamura and Yusuke Miyao, Generating Market Comments Referring to External Resources, In *Proceedings of The 11th International Conference on Natural Language Generation (INLG2018)*, pp. 135–139, 2018
- Tatsuya Ishigaki, Hiroya Takamura, and Manabu Okumura, Summarizing Lengthy Questions, In *Proceedings of The 8th International Joint Conference on Natural Language Processing (IJCNLP2017)*, pp. 792–800, 2017

国内会議論文

- 石垣 達也, Huang Hen-Hsen, Chen Hsin-Hsi, 高村 大也, 奥村 学, コピー機構を用いたクエリ指向ニューラル生成型要約, 情報処理学会第 240 回自然言語処理研究会, pp. 1–5, 2019
- 石垣 達也, 上垣外 英剛, 高村 大也, 奥村 学, 談話構造を考慮する階層的注意機構による抽出型ニューラル単一文書要約, 言語処理学会第 25 回年次大会, pp. 1555–1558, 2019

- 磯 颯, 上原 由衣, 石垣 達也, 能地 宏, 荒牧 英治, 小林 一郎, 宮尾 祐介, 岡崎 直観, 高村 大也, Data-to-Text における主題遷移のモデル化, 言語処理学会第 25 回年次大会, pp. 727–730, 2019
- 石垣 達也, 町田 和哉, 小林 隼人, 高村 大也, 奥村 学, Distant Supervision による質問要約, 情報処理学会第 236 回自然言語処理研究会, pp. 1–5, 2018
- 石垣 達也, 高村 大也, 奥村 学, 「長文質問」のための抽出型及び生成型要約, 情報処理学会第 232 回自然言語処理研究会, pp. 1–8, 2017

受賞

- **優秀研究賞**, 石垣 達也, Huang Hen-Hsen, Chen Hsin-Hsi, 高村 大也, 奥村 学, コピー機構を用いたクエリ指向ニューラル生成型要約, 情報処理学会第 240 回自然言語処理研究会, 2019
- **若手奨励賞**, 石垣 達也, 談話構造を考慮する階層的注意機構による抽出型ニューラル単一文書要約, 言語処理学会第 25 回年次大会, 2019

謝辞

本学位論文は著者が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に在籍中の成果をまとめたものです．本論文の審査を引き受けてくださった知能システム科学専攻の中本高道教授，小野功准教授，長谷川晶一准教授に感謝申し上げます．長い学位審査のの各段階において俯瞰的な視野からアドバイスを頂き，研究を加速させることができました．

本研究の遂行に当たり，指導教員である高村大也教授，奥村学教授，上垣外英剛助教にご指導を賜りました．感謝申し上げます．高村先生は，研究成果をなかなか出せなかった私を修士課程から 5 年間に渡り辛抱強く指導してくださいました．“一撃必殺の論文” を将来世に出せるよう，さらに精進したいと思います．奥村先生からは在学中から現在に至るまで，多くの貴重な機会に挑戦させて頂き，感謝しております．私にとって，2019 年 9 月から 3ヶ月間を過ごした国立台湾大学での生活はかけがえのないものになっています．これまで頂いたご恩を今後研究室に返せるよう，一層研究に勤しみたいと思います．上垣外さんには，プログラムの実装を何度も助けて頂き大変お世話になりました．上垣外さんの力なくてはこの論文は完成してません．

研究室の皆さまには大変お世話になりました．感謝申し上げます．とくに，研究室の快適な計算機環境を整備してくださった Admin 係の皆さまおかげで，数日かかる計算も途切れずに実験することが出来ました．休日返上で計算機環境を整備して頂き，ありがとうございました．研究室秘書の飯山信子さんには事務処理を一手に引き受けてくださり，事務処理の苦手な私にとってかけがえのない存在でした．改めて感謝申し上げます．

国立台湾大学に訪問研究学生として受け入れてくださり，いつも優しく的確な指導をしてくださった Hsin-Hsi Chen 教授に感謝いたします．国立台湾大学の Hen-Hsen Huang さん，Chung-Chi Chen さん，Min-Huang Fu さんには研究に関する議論だけでなく，台湾の食文化についても熱心に教えて頂きました．また台北を訪れた際には熱炒に行きましょう．

共同研究で多くのご指導を頂いた NTT コミュニケーション科学基礎研究所の平尾努さん，Yahoo! JAPAN 研究所の小林隼人さんに感謝申し上げます．要約研究の第一線で活躍されているお二方とともに進める研究は，スピード感があり，大きな刺激となりました．研究室 OB

で富士通研究所の森田一さん，牧野拓哉さんには事あるごとに進路や人間関係などの相談に乗って頂きました．学年が進み後輩が増えていく中，親身に相談に乗ってくださる先輩方の存在はとても大きなものでした．また，産業技術総合研究所においてテクニカルスタッフとして勤務中にお世話になった宮澤彬さん，青木花純さん，五島圭一さん，能地宏さん，牛久祥孝さん，宮尾祐介先生に感謝申し上げます．複数人が共同ですすめる産総研での研究は，研究室での活動とは違い新鮮で，とても勉強になりました．

東京工業大学グローバルリーダー教育院 (AGL) と先生方に感謝申し上げます．AGL には修士課程から 5 年間に渡り，分野横断的な講義に加え経済的な支援も頂きました．採択されていなければ，学位を取得することは難しかったと思います．特に山田圭介教授，大隅規由教授にはエンドユーザにまで製品として届ける道筋を考えながら技術開発をすることの重要性を繰り返しご指導いただきました．AGL で学んだことは学位研究にも，私の価値観にも影響を与えています．また，折に触れて食事などに付き合ってくれた AGL OB の千條浩史さん，加藤智裕さん，中村駿介さん，山本純也さんに感謝申し上げます．研究室にこもりがちな博士課程の生活において，専門分野の異なる方々との関わりを持つことが出来，大きな刺激になりました．

Jenny Judson さん，Neil Judson さん，Hao Xu さんはニュージーランド留学中のホストファミリーで，私が言語やコミュニケーションに関する技術に興味を持つきっかけを作ってくれました．多様な背景を持つ人々が集まるニュージーランドのような国で，計算機が人々の円滑なコミュニケーションをどのように手助けできるか，という問題は私の研究の大きな研究テーマの一つになっています．博士課程で研究が思うように進まず辛いときには，ロトルアの温かい人々や温泉からくる硫黄の香りを思い出し勇気づけられました．

茨城工業高等専門学校の小飼敬先生，滝沢陽三先生および白鴎大学の奥山慶洋先生は私の高専時代の指導教員であり，はじめての研究の指導をしてくださいました．世の中に無い技術を生み出す楽しさを教えてくださった最初の先生方であり，本論文の執筆の最初の一歩を踏み出すことが出来ました．改めて感謝申し上げます．

最後に家族に感謝申し上げます．女手一つで支えてくれた母には相当な苦労があったと思います．頭が上がりません．天国の父も一安心していることでしょう．修士課程のころから共に住み，精神的な支えであった妻に改めて感謝します．多くの方々に支えられ博士論文を完成させることができました．本当にありがとうございました．

令和元年 8 月 16 日