学修番号 16890515

修士論文

文書構造に着目したニューラル文書要約

小平 知範

2018 年 3 月 25 日

首都大学東京大学院
システムデザイン研究科 情報通信システム学域

小平 知範

審査委員：

小町 守 准教授　（主指導教員）

石川 博 教授　（副指導教員）

片山 薫 准教授　（副指導教員）

文書構造に着目したニューラル文書要約*

小平 知範

修論要旨

要約を構築する主な目的は，読み手が文書すべてを読むことなくその文書を理解できるようにすることである．特にニュース要約では，スマートフォンユーザは画面のサイズが限られているので，表示できる限られた量の要約を読みたい．これらの目的を達成するために，ポータブルデバイス向けの要約システムは重要な情報を含んだ要約を限られた要約長の中で生成しなければならない．

要約タスクには抽出型と抽象型の 2 つのアプローチがある．抽出型アプローチは要約を作るために文書の一部（文や句，単語など）を選ぶ．抽象型アプローチは文書に現れない単語も使って要約を生成する．抽出型アプローチ は元の文書から出力する表現を直接抽出するので，抽象型アプローチより文法的な要約を作ることができる．しかし，それでは元の文書に現れない単語を選ぶことができない．

抽象型要約は機械翻訳タスクとは異なり，おおよその出力は入力の文書から得ることができる．また，抽象型要約では主に Encoder-Decoder という機構を用いる．Encoder-Decoder モデルにおいて入力系列はソース，出力系列はターゲットと呼ばれる．Encoder-Decoder は，ソース（文書）の情報を読み取る RNN の Encoder と，その情報をもとにターゲット（要約）を生成していく Decoder を組み合わせたものである．入出力ともに系列の場合は sequence-to-sequence と呼ばれる．

Sequence-to-sequence を基に，要約中に入力の文書に現れない単語を含む抽象型文要約タスクに取り組まれている．CNN / Daily Mail データセットは様々な長さの文で構成された要約が含まれているので，構造化された要約を生成するために要約の構造情報の注釈を簡単につけることができない．そのため，彼らのモデルは構造的な要約の生成ができない．

そこで，本研究ではニュース要約のための構造的な要約（3 行要約）の生成に着目

*首都大学東京大学院 システムデザイン研究科 情報通信システム学域 修士論文, 学修番号 16890515, 2018 年 3 月 25 日.

し，我々は CNN / Daily Mail データセットと同量の要約データセットを Livedoor News から構築した．Livedoor News は 3 行要約とニュースを公開しているので，このデータセットを用いた解析は容易である．

3 行要約の生成を解析するために，我々はニューラルネットを用いたモデルを用いた．モデルを改善するために，我々は彼らのモデルを基に新しい機構を提案する．我々の貢献は以下である．

- データセットに対して，要約の構造の注釈付けと解析を行った．
- このデータセットの特徴を基に 3 行要約に適応したモデルを提案した．
- 3 行要約に着目した評価指標の提案をした．

本研究で作成したデータセット及び注釈付けされたデータセットは GitHub[1] にて公開した．

本論文の構成は以下のようになっている．第 1 章では本研究全体の概要，貢献を述べる．第 2 章では抽出型要約と抽象型要約についての関連研究について述べる．第 3 章ではニューラル要約の学習について述べる．第 4 章では大規模 3 行要約データセットの構築について詳しく述べる．第 5 章では 3 行要約の要約構造の分類モデルと 3 行要約の要約構造に適した fine-tuning について述べる．第 6 章では，要約を構造情報ごとに分類する実験結果について述べる．第 7 章では．要約の実験結果について述べる．第 8 章では，実験結果に対する考察を述べる．最後に第 9 章で本研究のまとめ，今後の展望について述べる．

[1] https://github.com/KodairaTomonori/ThreeLineSummaryDataset

Incorporating Document Structure into Neural Abstractive Summarization*

Tomonori Kodaira

Abstract

Neural network-based approaches have become widespread for abstractive text summarization. Previous models prevent repetition of the same contents in the summary, but do not explicitly take its information structure into account. One of the reasons they failed to model information structure of the generated summary is that the standard datasets, CNN / Daily Mail summarization tasks, include summaries of variable lengths. Thus, it is not clear how the first sentence contributes to the following sentences, and so forth. To address the lack of the dataset for structured summarization, we introduce a new dataset containing summaries consisting of only three bullet points, and propose a neural network-based abstractive summarization model considering information structure of the generated summary. Our contributions are as follows:

- We constructed a new summarization dataset, whose summaries are in the form of three sentences.
- We annotated and analyzed the structure of summaries in the dataset.
- Our model generates a summary considering the type of summary.

*Master's Thesis, Department of Information and Communication Systems, Graduate School of System Design, Tokyo Metropolitan University, Student ID 16890515, March 25, 2018.

目次

図目次

第 1 章　はじめに

要約を構築する主な目的は，読み手が文書すべてを読むことなくその文書を理解できるようにすることである．特に，スマートフォンユーザは画面のサイズが限られているので，表示できる限られた量の要約を読みたい．これらの目的を達成するために，ポータブルデバイス向けの要約システムは重要な情報を含んだ要約を限られた要約長の中で生成しなければならない．

要約タスクには抽出型と抽象型の 2 つのアプローチがある．抽出型アプローチは要約を作るために文書の一部（文や句，単語など）を選ぶ．抽象型アプローチは文書に現れない単語も使って要約を生成する．抽出型アプローチ [1, 2] は元の文書から出力する表現を直接抽出するので，抽象型アプローチより文法的な要約を作ることができる．しかし，それでは元の文書に現れない単語を選ぶことができない．

抽象型要約は機械翻訳タスクとは異なり，おおよその出力は入力の文書から得ることができる．また，抽象型要約では主に Encoder-Decoder という機構を用いる．Encoder-Decoder モデルにおいて入力系列はソース，出力系列はターゲットと呼ばれる．Encoder-Decoder は，ソース（文書）の情報を読み取る RNN の Encoder と，その情報をもとにターゲット（要約）を生成していく Decoder を組み合わせたものである．入出力ともに系列の場合は sequence-to-sequence と呼ばれる．

要約では，話の流れの一貫性を捉えるために 2 文間の意味的関係を表現する修辞構造理論 [3] が素性として用いられる．例えば，要約の 1 文目には基本的な情報，2 文目には 1 文目に対する追加情報が記述されているならば，2 文間の関係は “詳細”（Elaboration）に当たる．本研究ではこのような構造に着目して実験を行う．

Rush ら [4] は，Sutskever ら [5] が提案した sequence-to-sequence を基に，要約中に入力の文書に現れない単語を含む抽象型文要約タスクに取り組んだ．近年 Rush ら [4] の手法をもとに Nallapati らや See ら [6, 7] によってニューラルネットを用いた抽象型文書要約のアプローチが提案された．彼らの用いた CNN / Daily Mail データセットは様々な長さの文で構成された要約が含まれているので，構造化された要約を生成するために要約の構造情報の注釈を簡単につけることができない．そのため，彼らのモデルは構造的な要約の生成ができない．

そこで，本研究ではニュース要約のための構造的な要約（3 行要約）の生成に着

目し，本研究では CNN / Daily Mail データセットと同量の要約データセットを Livedoor News から構築した．Livedoor News は 3 行要約とニュースを公開しているので，このデータセットを用いた解析は容易である．

3 行要約の生成を解析するために，本研究では See ら [7] のモデルを用いた．本研究では彼らのモデルを基に構築したデータセットに特化したモデルを構築した．はじめに，構築したデータセットに対してアノテーションを少量行い構造情報の付与を行なった．次に，少量のデータセットを元に構造情報の自動付与を行なった．最後に，自動付与されたデータを用いて fine-tuning することにより，3 行要約に特化したモデルを作成した．また，システム要約の特徴を捉えるために新たに評価指標を提案した．

本研究の貢献は以下である．

- データセットに対して，要約の構造の注釈付けと解析を行った．
- このデータセットの特徴を基に 3 行要約に適応したモデルを提案した．
- 3 行要約に着目した評価指標の提案をした．

本研究で作成したデータセット及び注釈付けされたデータセットは GitHub[1] にて公開した．

本論文の構成は以下のようになっている．第 1 章では本研究全体の概要，貢献を述べる．第 2 章では抽出型要約と抽象型要約についての関連研究について述べる．第 3 章ではニューラル要約の学習について述べる．第 4 章では大規模 3 行要約データセットの構築について詳しく述べる．第 5 章では 3 行要約の要約構造の分類モデルと 3 行要約の要約構造に適した fine-tuning について述べる．第 6 章では，要約を構造情報ごとに分類する実験結果について述べる．第 7 章では．要約の実験結果について述べる．第 8 章では，実験結果に対する考察を述べる．最後に第 9 章で本研究のまとめ，今後の展望について述べる．

[1] https://github.com/KodairaTomonori/ThreeLineSummaryDataset

第 2 章　関連研究

この章では，抽出型要約と抽象型要約についての関連研究について述べる．

2.1　要約タスク

要約には，1 つの文書に対して 1 つの要約を生成する単一文書要約と複数の文書に対して 1 つの要約を生成する複数文書要約がある．単一文書要約はニュース記事のなどの大まかな概要をまとめるために用いられる．複数文書要約は複数の観点から書かれた記事や時系列のある記事等をまとめるために用いられる．

また，要約の種類としては，文書の内容を伝える報知的要約と，ある文書を読むべきか判断する材料としての指示的要約がある．

要約の作り方にも 2 種類あり，抽出型要約と抽象型要約がある．抽出型要約は文書中の文，句あるいは語を抜き出し並べ換えることで要約を作成する．抽象型要約は文書の情報を元に新たな文を作り出すことで要約を作成する．

本研究では 1 記事から 3 行からなる要約を生成する単一文書要約の抽象型要約に取り組む．

2.2　評価指標

要約の代表的な評価指標として ROUGE [8] スコアがある．ROUGE は正解要約とシステム要約間で単語の再現率を元にスコアを算出する．要約においては文書中の情報を伝えることが重要であるため，正解要約とシステム要約に対する n-gram の再現率を用いて要約の良さを測る ROUGE-N がある．これは以下の式で計算される．

$$ROUGE - N = \frac{\sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{gram_n \in S} Count(gram_n)} \tag{2.21}$$

ここで，S は正解要約，$gram_n$ は正解要約中に含まれる n-gram を示す．また，$Count_{match}(gram_n)$ はシステム要約と正解要約間で一致している n-gram の数を返す関数である．また, $Count(gram_n)$ は正解要約 n-gram に含まれている n-gram

の数を返す関数である．

次に ROUGE-L について説明する．二つの要約に対しての LCS（Longest Common Sequence）を算出し，より長ければ似ているという直感のもと作られた指標である．具体的には以下の式で計算される．

$$R_{lcs} = \frac{LCS(X,Y)}{m} \tag{2.22}$$

$$P_{lcs} = \frac{LCS(X,Y)}{n} \tag{2.23}$$

$$F_{lcs} = \frac{R_{lcs}P_{lcs}}{R_{lcs} + P_{lcs}} \tag{2.24}$$

ここで，X は正解要約，Y はシステム要約である．$LCS(X,Y)$ は二つの要約間の LCS の長さである．m は正解要約の長さ，n はシステム要約の長さを示している．再現率の計算には分母を m とすることで正解要約の内容をどれだけ出力できているかを示しており，適合率では分母を n とすることでシステム要約の内容がどれだけ正しいかを示している．

2.3 抽出型要約の関連研究

抽出型要約におけるベースラインとして用いられる手法として LEAD がある．LEAD は入力文書中の文を上から任意の数取ってくるものである．これは，文書中の重要な内容は文書の先頭にくるという仮定のもと用いられている．単純でありながら精度の高い手法である．

抽出型要約では ILP（整数計画法）を用いる手法 [2, 9, 10] がある．ILP では，最大化するスコアと制約が存在する．スコアは選ばれた要約に含まれる単語の異なり数や重要度などが用いられる．制約には，要約後の単語数や同じ単語を使う回数などが用いられる．このようにある制約のもとスコアを最大化するような文を抽出し，要約を作成する．

一例として，Hirao [10] らは単一文書要約をナップサック問題として定式化している．ナップサック問題では文書中の文に対して重要度を定義し，ある一定の要約長内で重要度が最大となるように要約を生成する．しかし，ナップサック問題では内容の冗長な表現などを選んでしまう場合がある．それに対処するために，彼らは

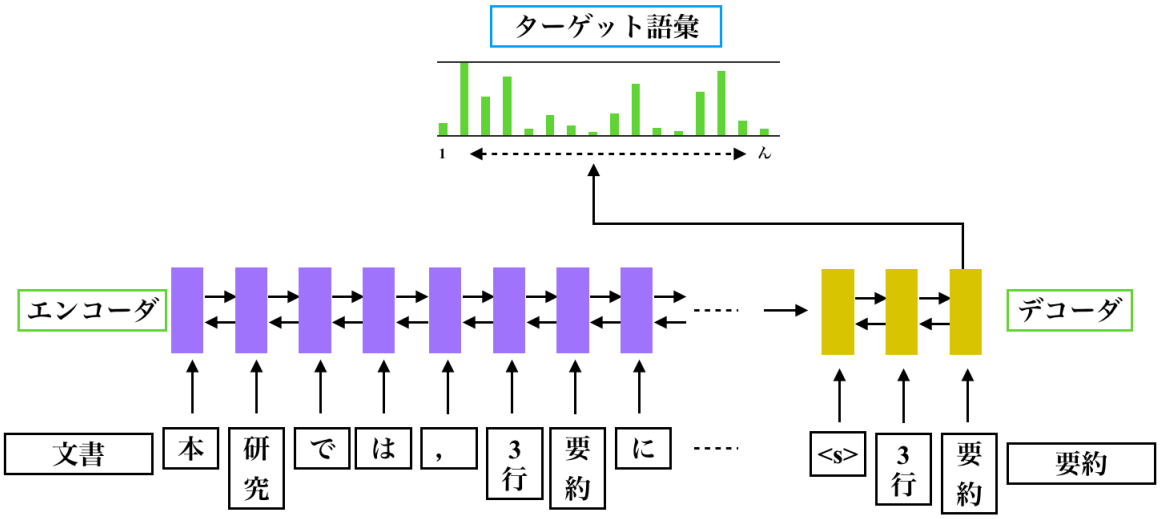

図 2.1　sequence-to-sequence の概略図.

ナップサック問題に冗長性の制約を設けた.

また，文書中の文や句を組み合わせて要約を作成するものもある．田中ら [11] は，文選択や文圧縮等を用い要約を作成している．彼らは LivedoorNews の記事と要約のペアを用いて要約の文字数，各文の文字数，文数に制約を加えている.

2.4　抽象型要約の関連研究

抽象型文要約では，Rush ら [4] が sequence-to-sequence モデル（図 2.4）を使った抽象型要約を生成する新しい要約手法を提案した．図 2.4 のように sequence-to-sequence はエンコーダ側で入力文書の情報を読み取り，デコーダ側において要約文の生成を行なっている．彼らは Gigaword コーパスと DUC-2004 において，世界最高精度を達成した．いずれのコーパスもニュース記事を含んでいるが，学習データの要約には修辞構造のようなものがないのでそれを考慮した出力にはなっていない.

CNN / Daily Mail 要約タスクの中の記事から複数文で構成されている要約を出力するタスクにおいて，抽象型文書要約に取り組んだ研究がある．Nallapati ら

[6] は Large Vocabulary Tric [12] や Switching Pointer-Generator，階層的ネットワークを Attention Encoder-Decoder モデルに取り入れ，このタスクにおいての改善モデルを提案した．Large Vocabulary Tric は要約側の単語のほとんどは入力文書からくるという特徴を利用し，ミニバッチごとに語彙を決めることで，低頻度の単語も含めた大きな語彙を使うことができる．ミニバッチとは，ミニバッチ学習における学習時の勾配の計算と重みの更新をするひとまとまりのことである．オンライン学習とは異なり 1 事例ごとに勾配を計算し重みを更新するのではなく，例えばミニバッチサイズが 32 の場合 32 事例の順伝播が終わった後，重みの更新を行う．ミニバッチに入る事例数は学習データからランダムに選ばれる．その際の損失関数の例を以下に示す．

$$L(t, x; w) = \frac{1}{n} \sum_{i=1}^{n} (t_i, x_i; w) \tag{2.41}$$

ここで，n はミニバッチサイズ，l は 1 事例に対する損失関数，x_i は学習データ，t_i は教師データ，w はパラメータである．Switching Pointer-Generator は Decoder が 1 つずつ単語を出力する各タイムステップで単語を生成するかソース側の単語をコピーするかを決める機構である．こうすることで，次に生成する単語が未知語の際に Pointer が選ばれることで，未知語の生成が可能になる．Attention Encoder-Decoder は Encoder-Decoder モデルでは単語生成時にソース側のどの単語に注目してるかを入力系列長の長さの確率分布として与える機構を付与したものである（4.1 節）．これは，ソース側とターゲット側の単語列には必ず対応関係があり，デコーダは単語生成時にそれらを明示的に与えるためである．階層的ネットワークは，文単位での要約に含めるかの情報量や生成している際にどの文に注目しているかを捉えるための文単位と単語単位の Attention を組み合わせたものである．彼らは複数文要約のための新しいデータセットを提案し，ベンチマークを構築した．この研究における出力は複数文であるが，出力は必ずしも文書構造を考慮したものではなく，出力も詳細に分析されていない．

See ら [7] は CNN / Daily Mail の要約タスクにおいて世界最高精度を達成した．彼らは Nallapati [6] らのモデルをベースに，Hybrid Pointer-Generator ネットワークと Coverage 機構を提案した．Hybrid Pointer-Generator は Attention と語彙の確率分布を重ね合わせて出力単語を決定するものである．2 つの分布を重

ね合わせることで未知語を考慮でき，なおかつソース側の単語を出力しやすくなる．Switching Pointer-Generator は Pointer が選ばれた際に確実にソース側から単語を選ぶことができる．Hybrid Pointer-Generator は同時に Decoder から出力された単語生成確率と Attention によるソース側の単語の確率分布を同時に考慮することができる．Coverage 機構はニューラルネットワークを用いた言語生成に特有の，同じ内容の言語表現を繰り返し生成してしまう問題を解決するためのものである．我々のモデルは彼らのモデルをベースとしているため，詳細な内容は 4 節にて示す．

第 3 章　ニューラル文書要約モデル

本研究のモデルは Attention Encoder-Decoder と Hybrid Pointer-Generator, Coverage 機構を組み合わせた See ら [7] のモデルをもとに構築した．まず，これらの機構について説明を行う．次に 3 行要約のための提案手法を説明する．

3.1　Attention Encoder-Decoder

入力単語列は記事のトークン，出力単語列は要約のトークンである．Encoder 側は 1 層の双方向 LSTM [13] を用い，Decoder 側には順方向の LSTM を用いる．Encoder によって生成される Encoder の隠れ層を h_i とする．それぞれのステップ t で，Decoder は出力の前単語の単語埋め込みベクトルと Decoder の状態 s^t を渡す．単語埋め込みベクトルは学習時には正解の要約を前単語として用い，テスト時には Decoder によって生成された前単語を用いる．Attention の分布 a^t は Bahdanau ら [14] と同様に以下のように計算される．

$$e_i^t = v^T \tanh(W_h h_i + W_s s^t + b_a) \tag{3.11}$$
$$a^t = \mathrm{softmax}(e^t) \tag{3.12}$$

ここで v は重みベクトルであり，W_h と W_s は重み行列，b_a はバイアスベクトルであり，それぞれ学習可能パラメータである．Attention の分布はタイムステップ t における Encoder の隠れ層の重要度を示す確率分布として表される．文脈ベクトル h_*^t は以下によって計算される．

$$h_*^t = \sum_i a_i^t h_i \tag{3.13}$$

この文脈ベクトルを Decoder の状態 s^t と連結し，2 つの線形変換を用い語彙分布 P_{vocab} が以下の計算で生成される．

$$P_{vocab} = \mathrm{softmax}(V'(V([s^t, h_*^t] + b) + b')) \tag{3.14}$$

ここで V と V' は重み行列，b と b' はバイアスベクトルであり，それぞれ学習可能なパラメータである．得られた P_{vocab} の確率分布から一番確率の高い単語 w^t が出力単語として選ばれる．各タイムステップ t におけるロスはターゲット側の単語 w_*^t の負の対数尤度で以下のように計算される．

$$\text{loss}_t = -\log P_{vocab}(w_*^t) \tag{3.15}$$

また，全体の系列に対してのロスは以下である．

$$\text{loss} = \frac{1}{T}\sum_{t=0}^{T} \text{loss}_t \tag{3.16}$$

ここで，T はシステム出力の単語数である．

3.2 Hybrid Pointer-Generator Network

Hybrid Pointer-Generator は See ら [7] によって提案され，Attention と語彙分布を組み合わせるものである．彼らの Pointer-Generator は Attention モデル（4.1 節）と Pointer Network [15] を組み合わせたものである．これはソース側の単語分布をターゲット側の単語分布と同様に考慮する．こうすることで Switching Pointer-Generator 同様，ソース側にある見たことのない単語を考慮することができ，未知語の生成問題に対応できる．さらに，ソース側の単語の生成確率が高くなり，ソース側と同じ単語を使うことが多い要約タスクでは有用である．

各タイムステップ t で Pointer-Generator モデルの生成確率 $p_{gen} \in [0, 1]$ は文脈ベクトル h_*^t と Decoder の状態 s^t，Decoder の入力である単語埋め込みベクトル x_t によって計算される．

$$p_{gen} = \sigma(v_{h^*}^T h_*^t + v_s^T s^t + v_x^T x_t + b_g) \tag{3.21}$$

ここでベクトル v_{h^*} と v_s，v_x は重みベクトル，b_g はスカラーのバイアスであり，それぞれ学習可能パラメータである，σ はシグモイド関数である．p_{gen} は単語分布

P_{vocab} か Attention の分布 a^t のどちらを用いるかの soft switch として用いる．各文書において，これらは拡張語彙を作り，それはターゲットの語彙とソースの語彙との和集合である．拡張語彙の生成確率は以下で計算される．

$$P(w) = p_{gen}P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i=w} a_i^t \tag{3.22}$$

もし w が out-of-vocabulary（OOV）ならば，$P_{vocab}(w)$ は 0 であり，また w がソース側の単語に存在しなければ $\sum_{i:w_i=w} a_i^t$ は 0 である．

3.3 Coverage Mechanism

See ら [7] は Encoder-Decoder モデルにおける繰り返しの問題を解決するために Coverage モデル [16] を改善させた．彼らのモデルでは，各 Decoder のタイムステップまでの Attention の分布の合計が Coverage ベクトル c^t として保存される．

$$c^t = \sum_{t'=0}^{t-1} a^{t'} \tag{3.31}$$

c^t はタイムステップ t までにそれぞれの単語に対してどれだけ注目したかを示す．ソース文書の単語に対する分布である coverage ベクトルは以下のように用いられる．

$$e_i^t = v^T \tanh(W_h h_i + W_s s^t + w_c c_i^t + b_a) \tag{3.32}$$

ここで，w_c は重みベクトルであり，学習可能パラメータである．彼らは同じ場所への繰り返しの Attention に対してペナルティを与える目的で，Coverage のロスを取り入れた新しいロス関数を構築した．

$$\text{loss}_t = -\log P(w_t^*) + \lambda \sum_i \min(a_i^t, c_i^t) \tag{3.33}$$

λ は同じ場所への繰り返しの Attention をどれだけ許容するかのパラメータである．これにより，$w_c c_i^t$ では Attention が同じ場所を繰り返し指すことを防ぐので，同じ内容を出力することを防ぐことに繋がる．

第 4 章　要約データセットの構築

本研究では田中ら [11] 同様 Livedoor News [1] から日本語の記事と要約のペアを収集した．この要約は人間の編集者によって書かれており，3 文で構成されている．詳細は後に示す．本研究は 2014 年 1 月から 2016 年 12 月までの期間でデータの収集を行い，得られた記事と要約は計 215,560 ペアとなった．収集したデータを分割し，トレーニングデータとして 213,160 ペア，検証データとして 1,200 ペア，テストデータとして 1,200 ペアとした．検証データとテストデータは 2016 年 1 月から 2016 年 12 月の期間のものから毎月 100 件ずつ抽出した．

4.1　記事の特徴

実際の記事と要約例を図 4.1[2]図 4.2[3]に示す．

それぞれの記事に対して，9 つのカテゴリー（国内，海外，IT 経済，芸能，スポーツ，映画，グルメ，女子，トレンド）から 1 つのカテゴリーが選ばれ，そのカテゴリーに対するいくつかのサブカテゴリーから 1 つのサブカテゴリーが選ばれている．さらに，特定のタグ（キーワードやキーフレーズ，より詳細なカテゴリー）が付与されている．収集したデータはニュースの記事とタイトル，抽象型要約とより短いタイトルが存在する．

このデータセットは上記のような多くの有用な情報を持つが，本研究では記事と要約のみを用いる．

---

[1] http://news.livedoor.com/

[2] http://news.livedoor.com/article/detail/14143155/（2018 年 1 月 11 日閲覧）

[3] http://news.livedoor.com/topics/detail/14143155/（2018 年 1 月 11 日閲覧）

[4] http://news.livedoor.com/topics/detail/12252068/（2018 年 1 月 11 日閲覧）, 記事を付録 A.1 に示す.

[5] http://news.livedoor.com/topics/detail/12244553/（2018 年 1 月 11 日閲覧）, 記事を付録 A.2 に示す.

[6] http://news.livedoor.com/topics/detail/12302174/（2018 年 1 月 11 日閲覧）, 記事を付録 A.3 に示す.

[7] http://news.livedoor.com/topics/detail/11098552/（2018 年 1 月 11 日閲覧）, 記事を付録 A.4 に示す.

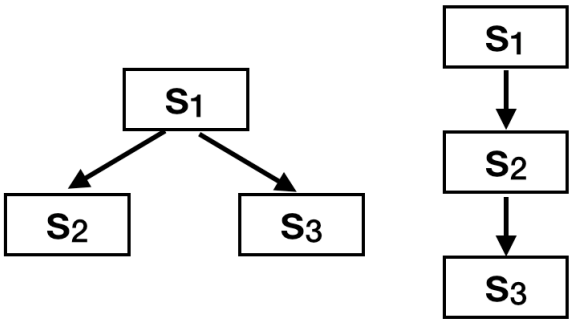

図 4.3　文書構造. 左: 並列タイプ. 右: 直列タイプ.

並列[4]	マクドナルド HP の、各国の違いを紹介している 日本はアメリカと似ているが、より情報を多く載せようという意図がみえる ドイツはバランスよく整理整頓され、フランスはモダンさが感じとれるという
直列 [5]	ソニーの VR ゴーグルが即完売し、生産が追いつかない人気ぶりだという 米投資銀行は 25 年に、VR・AR 関連の世界市場が約 9 兆 5000 億円になると予測 10 兆円市場はコンビニ全体の売上高と同規模で、投資家も注目しているそう

表 4.1　“並列” と “直列” タイプの要約例.

列挙型並列 [6]	コンビニ 3 社のボジョレー・ヌーヴォーを飲み比べている セブン-イレブンは多少渋味が強く、ファミリーマートは多少酸味が強いそう ローソンは後味がスッキリし人気が高く、予約分が完売した店舗もあるという
文分割型直列 [7]	山本昌氏が 23 日の「ジョブチューン」で星野仙一氏に殴られた話を明かした 「投げ終わってベンチ裏に来いと言われて多少かわいがられまして」と暴露 「顔が腫れ過ぎて降板した」と驚きの事実を伝えた

表 4.2　列挙型並列と文分割型直列タイプの要約例.

4.2　3 行要約に対する文書構造アノテーション

Livedoor News の要約は 3 文で構成されているため，出力の構造の解析が容易である．そこで，本研究では要約の一部である検証データとテストデータに対して要約の文書構造に対して注釈付けを行った.

それぞれの要約に対して話の流れに対応する一つのタグを付与した．多くの要約は**並列**タイプと**直列**タイプの 2 つの種類に分けられる（図 4.3）．“並列” と “直列”

タイプの例を表 4.1 に示す．最初の 2 文は 2 種類とも特徴が似ており，1 文目では主な出来事について記載され，2 文目は 1 文目に対する追加情報が記述されている．“並列” タイプは 3 文目が 2 文目とは異なる 1 文目に対する追加情報が書かれている．一方，“直列” タイプは 3 文目が 2 文目に対する追加情報が書かれている．つまり，“並列” タイプは 2 文目と 3 文目には特に順序はなく，“直列” タイプは 2 文目と 3 文目は順序には順序がある．

2 つのタグをアノテーションする中で，特徴的な構造をしている “列挙型並列”, “文分割型直列” を追加して，最終的には 4 タイプに分けた．追加したタグの例をそれぞれ表 4.2 に示す．列挙はあるものを紹介する時に要約の中に含まれることが多い．文分割は元々の文が長い場合に要約の中に現れる．これらは，主にスマートフォンで閲覧されることを想定してコンパクトに情報を提示する必要がある Livedoor News に特徴的な要約の例である．

4.3　アノテーションの結果と分析

アノテーションの結果を表 4.3 に示す．表の上部は “並列” と “直列” の 2 種類のタグのみでアノテーションした結果であり，下部は 4 種類のタグでアノテーションした結果である．表 4.3 に示すように，検証データとテストデータのいずれにおいても，約 70% の要約は “並列”，残りは “直列” のタグが振られる結果になった．表 4.3 では “列挙型並列” にタグづけされた中には単に例を並べるだけの文ではないものが存在した．

“並列” と “直列” に共通する特徴として，最初の文は主な出来事が記されている．おおよそ 2 文目は 1 文目の内容に対して結果の説明，詳細な情報，例などが書かれている．

一方，このデータセットでは 3 文目は様々な役割をしている．“並列” にタグ付けされた要約では，3 文目は 1 文目に依存している．一方で “直列” にタグ付けされた要約では，3 文目の内容は 2 文目に依存している．つまり，要約システムは 3 文目を生成する際にタグによって 1 文目か 2 文目のどちらに注意を向けるか決めなければならない．

本研究ではアノテーションする中で追加した “列挙型並列” と “文分割型直列” と

	検証	テスト	全て
並列	912	876	1,788
直列	288	324	612
並列	836	808	1,644
列挙型並列	76	68	144
直列	278	320	598
文分割型直列	10	4	14

表 4.3　3 行要約に対する文書構造アノテーションの結果.

なるデータは少量であるため，これらをそれぞれを “並列” 及び “直列” とみなし実験を行う（表 4.3）.

本節で行なったアノテーション結果を元に，次章では要約構造付きの検証データを正解データとし，トレーニングデータに対して要約構造自動分類を行う.

図 4.1　実際の記事例

図 4.2　実際の 3 行要約例.

第 5 章　要約構造分類モデル

本章では，4 章において作成及びアノテーションした要約構造データを用いた fine-tuning 用の文書要約データの作成について述べる．また，2 つのタイプそれぞれのデータを用いて fine-tuning するため，2 つのモデルどちらを出力するかを判定するための自動分類器の作成も行う．要約構造を捉えるために “並列” と “直列” タイプのデータを fine-tuning でそれぞれ使用する．しかし，タイプごとのデータ量は少量であり，学習に用いるには足りていない．また，学習用に大量のデータに対してアノテーションを行うにはコストが高い．そこで，本研究ではアノテーションした少量のデータを学習データとして用い，タグがついてないデータに対して自動タグづけを行う．

5.1　要約構造分類モデル

ここでは，与えられた要約に対して “並列” または “直列” ラベルを推定するモデルについて説明を行う．要約の単語列を x_i，出力ラベルを l とする．本研究では要約の情報を捉えるために，双方向 LSTM をエンコーダとして用いる．エンコードすることによって得られるエンコーダの隠れ状態の系列を h_i とする．順方向の LSTM の最後の隠れ層（$h_n^{forward}$）と逆方向の LSTM の最後の隠れ層（$h_1^{backward}$）を連結させたものを入力系列の情報として持つベクトル h を作成する．

$$h = [h_n^{forward}, h_1^{backward}] \tag{5.11}$$

ここで作られたベクトルに対して 2 種類の線形変換を適用し，2 つのラベルそれぞれに 2 次元のベクトルを構築する．(図 5.1).

$$y_{parallel} = \mathrm{softmax}(W_p h + b_p) \tag{5.12}$$
$$y_{sequence} = \mathrm{softmax}(W_s h + b_s) \tag{5.13}$$

ここで，$y_{parallel}$ と $y_{sequence}$ はそれぞれ並列と直列に対する 2 次元のベクトルで，1 次元目は対応するラベルでない確率，2 次元目は対応するラベルである確率を表す．W_p と W_s は重み行列であり，b_p と b_s は 2 次元のバイアスベクトルである．

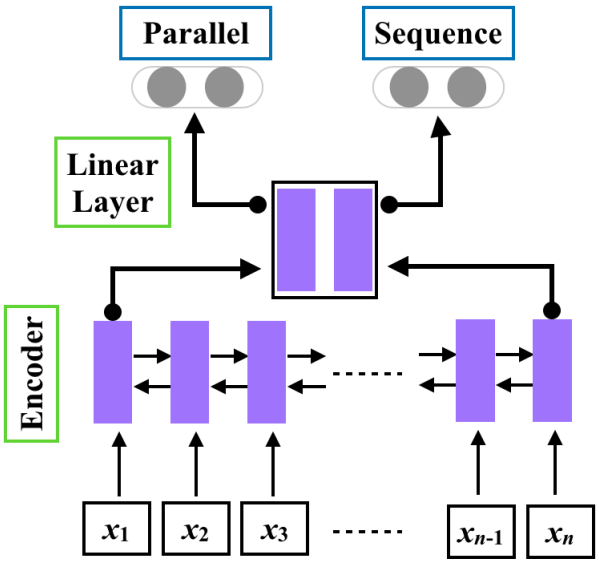

図 5.1　分類モデルの説明図.

5.2　要約構造に適応させる fine-tuning

3 章で先行研究の要約モデルの説明を行った．先行研究で用いられたデータセットでは，出力の文数が様々で要約の文書構造に着目したアプローチがあまりない．しかし，本研究で構築したデータセットでは要約が 3 文であることと，要約の文書構造が 2 種類であることが 4 章で明らかになっている．

先行研究の要約モデルを学習したモデルは “並列” と “直列” のどちらのタイプの要約を出力すれば良いかという情報を受け取らずに学習をしている．そこで，トレーニング済みのモデルに対して fine-tuning を適用することでそれぞれのタイプに適応したモデルの構築を行う．

はじめに，全てのトレーニングデータを用いてモデルの学習を行う．次に各タイプに適したモデルを構築するために fine-tuning を行う．4 章で説明した要約分類モデルで推定された各タイプのデータを用い，タイプごとに追加学習を行う．

5.3　記事を入力とした自動分類モデル

fine-tuning により構築した 2 つのモデルのどちらを要約に用いるかを決めるため，同様のネットワークを用いて分類モデルの作成を行う．テストデータを分類する際には，正解である要約を入力として用いることができないため，記事を入力データとする．ここで構築されたモデルから得られた分類結果を元にシステム要約を出力する．

第 6 章　要約構造分類実験

本章では 4 章で説明を行った要約の自動分類の実験について述べる．始めに実験設定を説明し，その後実験結果について述べる．

6.1　要約を入力とした実験設定

トレーニングデータとして要約の学習に用いる検証データの 1,020 件，検証データとして検証データの残りの 180 件，分類対象は要約の学習に用いるトレーニングデータを使用する．

記事と要約の単語分割には 形態素解析エンジン MeCab v0.996 [1]を用いる．辞書には IPAdic（v2.7.0）を使用する．LSTM の隠れ状態の次元数を 256 次元，単語埋め込みベクトルを 256 次元に設定する．語彙サイズは 2,350 であり，これらは頻度 2 以上のものを選択している．モデルの学習時に Adagrad [17] を学習率 0.01 で用いた．

学習に用いるデータは “並列” タイプが大半を占めるため，アンダーサンプリングを適用した．学習では小さなデータを用いるため，閾値を設け確度の高いラベルの獲得を行う．各ラベルの分類の適合率が 0.8 以上になるまで閾値を調整した．

また，5 章で説明した分類モデルを用いてテストデータを事前に 2 つのタイプへの分類を行なった．分類するデータはテストデータであるため，記事を入力，2 つのタイプを推定する．6 章同様のパラメータ設定でモデルの構築を行い，テストデータを分類した．

6.2　実験結果

各ラベルにおける検証データの精度と要約モデルのトレーニングデータに対して分類した結果獲得したデータ数を表 6.1 に示す．今回は適合率を高く設定したため，再現率は低くなっている．並列タイプと直列タイプに分類された要約はそれぞ

[1]https://github.com/taku910/mecab

	適合率	再現率	ラベル付けされた数
並列	0.897	0.325	53,809
直列	0.833	0.156	7,813

表 6.1　2 種類のラベルの分類結果.

	適合率	再現率	F 値
並列	0.71	0.53	0.61
直列	0.25	0.42	0.31

表 6.2　2 種類のラベルの分類結果.

れ 53,809 件，7,813 件得られた．本研究では並列タイプのデータ量が多い結果となったが，比較のため直列のデータ数と同数だけを用いる．

6.3　記事を入力とした実験結果

テストデータにおいて記事を入力として要約構造分類実験をした結果を表 6.2 に示す．2 つのラベルの正解率は 0.50 であった．直列タイプにおける適合率，再現率が低くなっている．

第 7 章　3 行要約実験

7.1　実験設定

3 行要約実験では 3 節で説明した See ら [7] のモデルを用いた．実験においては，See ら [7] 同様隠れ層を 256 次元に，単語埋め込みベクトルを 128 次元に設定した．語彙サイズはソース側とターゲット側それぞれ 50,000 とした．

記事の頭から 400 単語のみを使用し，トレーニング及びテストを行った．これは，先行研究において重要な内容が先頭付近にあるという仮説のもと行った．また，デコード時には 70 単語以上生成する場合は 70 単語で生成を止めた．モデルの学習時には Adagrad [17] を学習率 0.15 で使用し，グラディエントクリッピング [18] の最大グラディエント L2 ノルムを 2.0 とした．検証データを用いて，100 イテレーションごとに ROUGE-L で評価を行い，スコアが最大となるモデルを最終的なモデルとした．

fine-tuning では，6 章で得られた 2 つのデータを用いた．すべてのトレーニングデータで学習した先行研究のモデルをそれぞれのタイプに合わせて追加学習を行った．

また，6.3 節で行なったテストデータを事前に 2 つのタイプへの分類した結果を用い，2 つのタイプのどちらのモデルを用いるかを決めるモデル（$Merge$）も作成し実験を行う．

7.2　評価方法

評価するモデルは 7.1 で述べた 3 つのモデルである．本研究ではシステム出力の要約の評価指標として ROUGE-1 と ROUGE-2, ROUGE-L の F_1 スコアの平均を用いる．評価対象はすべてのテストデータ（All），“並列” タイプのみ（$Parallel$），“直列” タイプのみ（$Sequence$）の 3 つである．さらに，システム出力の各文を評価するために正解要約の 3 文と 1 対 1 のペアを作る．ここで作られるペアは 3 つのペアの ROUGE-L スコアの平均が最大となり，システムの各文が異なる正解要約内の 1 文と対応する．

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
All	47.22	21.65	32.96	47.41	21.75	33.36	**47.88**	**21.92**	33.40	47.69	21.75	**33.41**
Parallel	47.02	21.83	32.92	47.16	21.94	33.21	47.43	21.97	33.20	**47.83**	**22.00**	**33.63**
Sequence	47.80	21.20	33.05	48.14	21.34	**33.85**	**49.11**	**21.86**	33.80	47.34	21.09	32.80

表 7.1　各データでの評価結果.

7.3　実験結果

本節では 3 つのモデルを ROUGE [8] スコアで評価した結果を述べる．7.3.1 節ではシステム出力と正解要約全体に対して ROUGE-1, ROUGE-2, ROUGE-L を計算した結果を示す．7.3.2 節ではシステム出力と正解要約それぞれの文を順番に評価した結果について示す．7.3.3 節ではシステム出力のそれぞれの文に対して，ROUGE-L が最大となるペアを作り，それらのペアごとの評価結果を示す.

7.3.2,7.3.3 節では $1st$，$2nd$ 及び $3rd$ 行はシステム要約の各文における評価である．Ave の行は $1st$ と $2nd$，$3rd$ のスコアの平均を示している.

7.3.1　ROUGE による評価結果

表 7.1 に ROUGE による評価結果を示す.

表の *All* について述べる．表 7.5 において，提案モデルがベースラインのモデルより高い ROUGE スコアを達成している．$Parallel$ では，すべての提案モデルがベースラインより高いスコアとなった．$Sequence$ においては $Sequence\_Train$ モデルがベースラインのモデルを大きく上回っている（ROUGE-1 +2.31, ROUGE-2 +0.66, ROUGE-L +0.75）.

7.3.2　各文に対する評価結果

順番にシステム要約と正解要約の文でペアを作り，評価した結果を表 7.2, 7.3, 7.4 に示す．テストデータ全体で評価した際には $Merge$ が $3rd$ を除いてより高いスコアとなった（表 7.2）．また，“Parallel” タイプのテストデータにおいても同様の傾

向が見られる（表 7.3）．“Sequence” タイプのテストデータでは $Parallel\_Train$ が多くの評価指標で高いスコアとなった．

7.3.3 ROUGE-L を用いてペアを作成し，各文に対する評価結果

ROUGE-L を用いてシステム要約の 3 文のスコアの平均が最大となるようなペアを作成し，評価を行なった結果を表 7.5, 7.6, 7.7 に示す．

$1st$ の結果について述べる．この列は $Merge$ のモデルのスコアが優位である．表 7.7 ではベースラインのモデルと比べると，ROUGE-1, ROUGE-L ともに 1.00 近く上回っている．

続いて，$2nd$ の結果について述べる．$Sequence\_Train$ が総合的に見て高いスコアを出している．表 7.7 においてはベースラインのモデルに比べ，ROUGE-1 は 3.13, ROUGE-2 は 3.24, ROUGE-L は 2.21 上回っている．$Merge$ は表 7.6 でベースラインのモデルより 1.00 近く低いスコアが出ていることがわかる．

$3rd$ の結果について述べる．表 7.5 7.6 では他の 2 つのモデルと比べ高いスコアを達成している．表 7.7 では今までの結果とは異なり，ベースラインのモデルにおいて高いスコアが出ている．

最後に Ave の結果について述べる．3 文それぞれの平均を取ると，$Sequence\_Train$ がどのデータでも一番高いスコアを達成している．また，$Parallel\_Train$ においても，ベースラインのモデルを上回る結果となった．

全体的な傾向としては，1 文目から 3 文目に進むにつれて全体的に ROUGE スコアが下がる傾向にある．また，fine-tuning を行ったモデルはベースラインのモデルより高い結果となることが多い．

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
1st	47.04	27.12	34.01	48.01	27.86	34.54	47.81	27.27	33.95	**48.90**	**28.32**	**35.12**
2nd	28.99	9.84	18.71	29.27	9.84	18.78	**29.59**	10.18	19.10	29.50	**10.38**	**19.18**
3rd	26.79	8.05	17.87	27.14	8.26	18.19	**27.58**	**8.59**	**18.42**	26.85	7.67	17.91
ave	34.27	15.00	23.53	34.81	15.32	23.83	35.00	15.35	23.83	**35.08**	**15.46**	**24.07**

表 7.2　1-1, 2-2, 3-3 でペアを作った，すべてのテストデータにおける評価結果.

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
1st	47.10	27.56	34.35	48.28	**28.60**	34.95	47.90	27.88	34.28	**49.00**	28.37	**35.25**
2nd	29.01	10.16	18.81	28.92	9.72	18.56	29.40	**10.28**	**19.08**	**29.43**	10.27	**19.08**
3rd	25.99	7.85	17.29	26.30	7.94	17.48	26.87	**8.58**	17.89	**27.02**	7.73	**18.00**
ave	34.03	15.19	23.48	34.50	15.42	23.66	34.73	**15.58**	23.75	**35.15**	15.46	**24.11**

表 7.3　1-1, 2-2, 3-3 でペアを作った，“並列” タイプのテストデータにおける評価結果

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
1st	46.88	25.93	33.10	47.27	25.86	33.42	47.56	25.61	33.09	**48.62**	**28.18**	**34.76**
2nd	28.94	8.98	18.45	**30.20**	10.17	19.38	30.12	9.90	19.17	29.68	**10.67**	**19.45**
3rd	28.94	8.57	19.45	29.41	**9.12**	**20.10**	**29.50**	8.63	19.86	26.38	7.51	17.64
ave	34.92	14.49	23.66	35.63	15.05	**24.30**	**35.72**	14.72	24.04	34.89	**15.45**	23.95

表 7.4　1-1, 2-2, 3-3 でペアを作った，“直列” タイプのテストデータにおける評価結果.

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
1st	48.23	28.42	35.29	48.86	28.80	35.53	49.20	28.84	35.38	**49.48**	**29.15**	**35.82**
2nd	34.37	14.50	23.38	34.30	14.18	23.28	**35.09**	**15.41**	**24.15**	34.69	15.24	23.78
3rd	31.91	12.36	22.00	**32.28**	**12.80**	**22.51**	32.01	12.18	21.97	31.67	11.46	21.80
ave	38.17	18.43	26.89	38.48	18.59	27.11	**38.77**	**18.81**	**27.17**	38.61	18.62	27.13

表 7.5　スコアが最大となるペアを作った，すべてのテストデータにおける評価結果.

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
1st	48.38	28.97	35.68	48.75	29.30	35.68	49.35	**29.38**	35.68	**49.63**	29.18	**35.95**
2nd	34.48	14.93	23.40	33.62	13.89	22.81	34.32	14.97	23.64	**34.51**	**15.13**	**23.74**
3rd	31.62	12.53	21.82	**32.45**	**13.29**	**22.67**	31.58	12.50	21.80	32.12	11.80	22.16
ave	38.16	18.81	26.97	38.27	18.82	27.05	38.42	**18.95**	27.04	**38.75**	18.70	**27.28**

表 7.6　スコアが最大となるペアを作った，“並列” タイプのテストデータにおける評価結果.

	Coverage			Parallel_Train			Sequence_Train			Merge		
	ROUGE			ROUGE			ROUGE			ROUGE		
	1	2	L	1	2	L	1	2	L	1	2	L
1st	47.83	26.92	34.25	**49.18**	27.45	35.14	48.81	27.38	34.54	49.09	**29.08**	**35.47**
2nd	34.05	13.34	23.32	36.14	14.96	24.55	**37.18**	**16.58**	**25.53**	35.17	15.52	23.90
3rd	32.68	**11.93**	**22.47**	31.82	11.48	22.08	**33.17**	11.34	22.42	30.46	10.55	20.84
ave	38.19	17.40	26.68	39.05	17.96	27.26	**39.72**	**18.43**	**27.50**	38.24	18.38	26.74

表 7.7　スコアが最大となるペアを作った，“直列” タイプのテストデータにおける評価結果.

第 8 章　考察

8.1　評価結果

7 章では，通常の評価方法（*All*）に加え，1 文ごとにペアを作り評価を行った．*All* のように評価をしてしまうと，文の対応が考慮できないため，*Ave* より高めのスコアが出ている．例えば，正解要約の 1 文目の bi-gram とシステム要約の 3 文目に bi-gram があったとしてもスコアが上がるため，*All* のスコアが高くなってしまう．文ごとに評価することで，何文目が生成しやすいまたはしにくいというのが理解できる．

表 7.5 では，1*st* は 4 章で述べた通り 2 つのタイプどちらも主な出来事が書かれているためスコアが高い．しかし，2，3 文目となると 1*st* に比べ 10.00 以上の差が開いている．2，3 文目は各タイプによって役割が違いモデルがどちらを出せばいいのか認識できないので，このような結果となっていると考えられる．しかし，これは Encoder-Decoder モデルにおける長文を生成する際の特徴でもあるため一概には言えない．

All と *Ave* の評価結果を見ると *Sequence_Train* が一番高い ROUGE スコアを出している．これはもともとのトレーニングデータには “並列” タイプの要約が多いと推測できる（表 4.3）ため，スコアの変化が大きいと考えられる．モデルを “直列” タイプで fine-tuning すると，順番に前の文を考慮しながら生成するようになるため，このような結果になったと考えられる．なぜなら，もともと “並列” タイプが多いトレーニングデータであるため，モデルは 3 文目生成時には 1 文目の内容を考慮しつつ，2 文目と被らない内容を出力しようとするため必要な情報量が多く，今までの生成内容を考慮しきれていないと考えられる．

8.2　各ペアごとの評価について

本節では，システム要約をより詳しく分析するために，ペアの組み合わせごとに分け，評価を行った．その結果を表 8.1 に示す．一番左の列はシステム要約の各文が何文目とペアになったかを示している．太字は行内において 1*st* と 2*nd*，3*rd* の

		1st			2nd			3rd		
ペア	ペアの数	ROUGE			ROUGE			ROUGE		
		1	2	L	1	2	L	1	2	L
123	528(44.0%)	55.07	34.56	40.22	36.93	17.31	25.21	**32.64**	**13.19**	**22.62**
132	381(31.8%)	52.38	31.94	37.72	33.59	14.14	23.15	**30.83**	**11.00**	**20.93**
213	100 (8.3%)	38.03	18.30	26.50	**31.32**	**10.34**	**21.25**	34.73	15.47	24.64
231	66 (5.5%)	41.46	21.95	28.09	37.22	17.95	27.23	**26.96**	**6.02**	**18.08**
312	64 (5.3%)	**28.37**	**7.78**	**19.26**	34.19	13.56	23.36	34.68	13.60	22.76
321	61 (5.1%)	**27.10**	**6.74**	**18.15**	33.36	14.36	23.53	32.16	10.67	21.85

表 8.1　各ペアの組み合わせごとの評価結果

中で**一番低い**スコアを示している．例えば，‘132’ の場合システム要約の 1 文目と正解要約の 1 文目，システム要約の 2 文目と正解要約の 3 文目，システム要約の 3 文目と正解要約の 2 文目がペアとなったことを示している．

結果を見ると ‘123’ が一番多くの割合を占めている．これは “並列” タイプと “直列” タイプの要約がうまく生成できた事例数だと考えられる．また，次に多い ‘132’ では “並列” タイプの事例が生成できた事例数だと考えられる．なぜなら，“並列” タイプの場合 2，3 文目は順不同であるためである．

また，特徴としては順番が異なる文が低いスコアを出している．例えば，‘213’ を見ると逆順となっているシステム要約の 2 文目と正解要約の 1 文目のペアが一番低いスコアを出している．これはシステム要約の一番目に主な内容が来るはずが，2 文目に来てしまっていると考えられる．また，‘312’，‘321’ では，最初の 2 文で本来の “並列” の 2，3 文目の内容を生成してしまっていると推測できる．

8.3　評価方法

本研究では 7 章で 3 通りの評価を行なった．7.3.1 節では，要約単位での評価，7.3.2 節では 3 行要約という特徴に基づき各文ごとでの評価，7.3.3 節では ROUGE-L のスコアに基づいて文ごとでの評価を行なった．要約単位の評価では，要約全体の良さを測ることができるが，各文に対する評価が行えないため，文ごとにペアを作成し，評価を行なった．各文ごとの評価を行なった方が，理解できることが多く，

有用性が高い．例えば，1 文目の評価結果が 2，3 文目より高く，文構造を捉えきれていないことが明らかになった．

7.3.2 節では単純にシステム要約と正解要約の各文で順番にペアを作成，評価を行ない，7.3.3 節では ROUGE-L のスコアを用いてペアを作成し評価を行なった．システム出力は必ずしも正解要約通りの順番で内容が構成されているとは限らないため，7.3.2 節における評価結果では，ROUGE-2 と ROUGE-L の評価が低い結果となっている．そのため，7.3.3 節で ROUGE-L のスコアを用いて各文に対応するペアを作成することで，適切な評価を行うことができた．また，各システム要約ごとに異なるペアを作成することにより，8.2 節においてシステムがどのような順番で文を出力しているかがわかるようになった．

8.4 分析

8.4.1 前文との関連性

出力例を表 8.2 に示す．この例の正解要約は，1 文目には日本人選手が所属するチームの試合結果，2 文目にはその選手の活躍内容とその賞賛，最後に選手の能力の評価を述べている．2，3 文目で 1 文目で登場する選手に対する賞賛の内容なので，“並列” タイプのタグをつけた例である．

3 つのシステム要約の 1 文目はどれも試合結果に関して述べられていない．しかし，この情報は記事中には記述されておらず，生成は難しい．*Paralel_Train* に関しては，選手を褒める内容ではなく，責める内容が書かれている．2 文目では，*Sequence_Train* のみ，正解要約と同等の内容が書かれている．3 文目では，3 システムともに同じ内容であるが，正解とは違う内容が出力されている．*Sequence_Train* の要約では，生成した 2 文目で記事の 2 文目の内容を生成できているのは，1 段落目の情報が繋がっていると捉えられていると考えられる．モデルに入力する際には，段落情報が含まれてはいない．*Sequence_Train* は “直列” タイプの前文の関連性を強く考慮するようになっているためだと考えられる．

出力例を表 8.3 に示す．この例の正解要約は，1 文目にある女優に対する話題，2 文目にその話題のエピソード，3 文目にそのエピソードの追加情報が書かれている．3 文目が 2 文目の追加情報となっているので，“直列” タイプのタグをつけた例で

ある．

3 つのシステム出力の 1 文目では，正解要約と同等の内容が書かれている．2，3 文目では $Sequence\_Train$ 以外の出力では，一貫性がなく，異なる情報を伝える要約となっている．$Sequence\_Train$ では，全体の話の流れと内容が正解要約と同等のものとなっている．これは，3 段落目の内容が関連のある内容であるので，2 文に分けて出力されているのがわかる．表 8.2 と同様に段落内の内容を続けて出力しているのがわかる．

上記の例のように，“直列” タイプのデータを用いて fine-tuning することによって，記事中の文間の繋がりを保つようなモデルとなっていることがわかる．

8.4.2　記事の入力長の制約

出力例を表 8.4 に示す．この例の正解要約は，1 文目に軽自動車の最新装備の話題，2 文目に最新の設備を搭載した軽自動車の情報，3 文目に 2 文目とは異なる自動車の情報が述べられている．正解要約で述べられている情報は記事の 5 段落以降の内容である．

すべてのシステム出力において，主な話題として高齢者に関連したことが出力されている．これは，トレーニング時に先頭 400 語のみを用いるという実験設定があるため，このように長い記事の場合，前半記事しか入力情報として出現しないからである．また，ニュース記事では前半部分で主な出来事が書かれていることが多いため，このような記事に対応するためには，前半部分が導入部分であり，後半部分に主要な情報が書かれていることを理解しなければならない．例えば，トレーニングデータの記事を用意する際にの前段階として重要な段落の抽出，またはトレーニングデータの記事中の文に対してトピックを割り当てる等の工程が必要となると考えられる．

8.5　注目箇所の誤り

出力例を表 8.5 に示す．この例の正解要約は，1 文目にアルバムの発売について，2 文目にそのアルバムの内容，3 文目にそのアルバムの特典について書かれている．

3 つのシステム要約の 1 文目は正解要約と同等の内容が出力されている．*Parallel_Train* の 2 文目ではアルバムの内容，3 文目では特典について書かれており，1 文目の追加情報として正しいものが確認できる．*Coverage* と *Sequence_Train* では 2，3 文目の出力はある店舗での限定的な情報を述べており，1 文目の追加情報として適していない内容となっている．*Parallel_Train* では，“並列” タイプで fine-tuning することにより，1 文目の追加情報が記載されている 2，3 段落目の主要な情報を抜き出せていると考えられる．*Sequence_Train* の出力では，2 文目に ‘TSUTAYA’ の先着特典，3 文目にはそこでのレンタルが開始される日が書かれているため，“直列” タイプの要約構造で書かれている．しかし，この記事での主要な事柄はアルバムの内容であるため不適切である．表 8.3 でも述べたように *Sequence_Train* では隣接する関連のある 2 文から情報を引き出してくる例が多く見られた．そのため，今回のような例では，適切な要約が生成できなかったと考えられる．

8.5.1 考察

分析の結果，*Sequence_Train* で学習することで，表 8.3 のように隣接する文から要約を生成できる例が多くあった．*Sequence_Train* は ROUGE におけるスコアも高く，3 行要約データセットではそのような正解要約が多いと考えられる．そのため，要約構造分類実験によって得られた “直列” タイプのデータはこのデータセットにおいて質の高いデータと言える．

エラー例からわかるように，記事の情報を読み取るには 400 単語という制約があり，いくつかのトピックが含まれる記事が存在するため，これらのような例では提案したモデルでの要約の生成は難しい．トレーニングデータ内の記事中の文に付加情報を追加したり，冗長な文を削除するなどの改善が必要である．または，本研究ではモデルをタイプごとに 2 つ用いたが，これらを記事ごとにどちらで生成するか

を自動で選択できるような機構が必要である.

[1]http://news.livedoor.com/article/detail/12016209/（2018 年 1 月 11 日閲覧）
[2]http://news.livedoor.com/topics/detail/12016209/ （2018 年 1 月 11 日閲覧）
[3]http://news.livedoor.com/article/detail/12402230/（2018 年 1 月 11 日閲覧）
[4]http://news.livedoor.com/topics/detail/12402230/ （2018 年 1 月 11 日閲覧）
[5]http://news.livedoor.com/article/detail/11159352/（2018 年 1 月 11 日閲覧）
[6]http://news.livedoor.com/topics/detail/11159352/ （2018 年 1 月 11 日閲覧）
[7]http://news.livedoor.com/article/detail/11155429/（2018 年 1 月 11 日閲覧）
[8]http://news.livedoor.com/topics/detail/11155429/ （2018 年 1 月 11 日閲覧）

記事 [1]	ケルンの地元紙エクスプレスは、先日のヴォルフスブルク戦で大迫勇也がその力を発揮したと賞賛。ヴォルフスブルク守備陣を翻弄してみせた日本代表ＦＷが、その力を発揮したと伝えた。 昨季の大迫は、決定力不足から熱いケルンファンからの厳しい声を浴びて、 シュテーガー監督が敢えてホーム戦での起用を避けるという事態にまで発展。 それでもシュテーガー監督やマネージャーのヨルグ・シュマッケ氏は、常に大迫のことを「素晴らしいサッカー選手だ」と擁護。 だが一方でＣＦモデストがチームトップの得点をマークする活躍を披露しており、「ＣＦを本職としていることはわかっていた」シュテーガー監督だったが、「どうしても起用するため」、結果的に様々ポジションで試行錯誤を繰り返す結果となっていた。 しかし今季のケルンは、開幕から４-４-２システムを採用しており、「今ははっきりとしている。彼はＣＦとしてプレーする」とシュテーガー監督は断言。ポカール１回戦でも２ゴールし、代表戦期間中のテストマッチでも連発。 そしてルドネフスの負傷でつかんだ今季初先発のチャンスで、大迫は”キープ力、スピード、視野の広さ、そして得点の脅威を試合を通じてみせていた”とエクスプレス紙が賞賛する活躍を披露し、”ケルンのＦW陣のなかで、全ての要素をもっとも兼ね備えた存在”とさえ評された。 ただ得点に関しては、この日はカスティールスの好セーブでお預けとなってしまったのだが、このまま本職のＣＦでこの好調を維持し、定位置確保取りに向けて突き進んでいきたい。
正解要約 [2]	大迫 勇 也 が 所属 する ケルン は 10 日 、 ヴォルフスブルク に ０ － ０ で 引き分け た 地元 紙 は 、 大迫 が ヴォルフスブルク 守備 陣 を 翻弄 し て みせ た と 称賛 ケルン の FW 陣 の なか で 、 全て の 要素 を もっとも 兼ね備え た 存在 と 評し た
Coverage	ケルン の 地元 紙 エクスプレス は 、 大迫 勇 也 が 力 を 発揮 し た と 伝え た 2015 年 の ヴォルフスブルク 戦 で 大迫 の こと を 「 素晴らしい サッカー 選手 だ 」 と 擁護 「 今 は はっきり と し て いる こと は わかっ て い た 」 と 語っ て い た
Parallel_Train	ケルン の 地元 紙 エクスプレス が 、 チーム トップ の 得点 ファン から の 厳しい 声 を 浴び た 「 ＣＦ を 本職 と し て いる こと は わかっ て い た 」 と 、 シュテーガー 監督 は 好 セーブ で お預け 昨季 の 大迫 勇 也 は 「 今 は はっきり と し て いる 。 彼 は ＣＦ として プレー する 」 と も
Sequence_Train	ケルン の 地元 紙 エクスプレス は 、 ヴォルフスブルク 戦 で 大迫 勇 也 が その 力 を 発揮 し た と 伝え た ヴォルフスブルク 守備 陣 を 翻弄 し て みせ た 日本 代表 ＦＷ が 、 その 力 を 発揮 し た と 賞賛 「 今 は はっきり と し て いる こと は わかっ て い た 」 と 擁護 し て い た

表 8.2　出力例１（並列）.

記事 [3]	10 日放送の「王様のブランチ」（TBS 系）で小池栄子が、本気で女優の道に進むきっかけとなった大物俳優の言葉を明かした。 番組では「ブランチタレント名鑑」のコーナーで、ゲストの小池が女優として大きなターニングポイントになった出来事を語った。小池は 2009 年にドラマ「スマイル」（同局）に新米弁護士役で出演していたとき、バラエティーと女優業の両立で悩んでいたという。 そんな小池に、共演していた中井貴一が「女優をやった方がいい」「いま新しく女優をやりたいなら、そっちに重きを置いて、何かを手放してやる覚悟も必要」だとアドバイスしてくれたというのだ。 この話を聞いた鈴木あきえが「中井貴一さんは、なぜ小池さんを見て『女優をやった方がいい』と、おっしゃったんですかね？」と質問。これに小池は「女優をやりたいんだけど、『バラエティー出身ですし』『グラビア出身ですし』と逃げる自分に、たぶん気づいていた。『そこを言い訳にすんな』とお叱りだったとも思う」と回想した。 一方、小池は「ブランチ」司会の谷原章介に対しては「特に何も教えてくれなかった」とクレームをつけ、笑いを誘っていた。
正解要約[4]	10 日 の 番組 で 小池 栄子 が 本気 で 女優 業 に 挑む こと に なっ た きっかけ を 明かし た ドラマ で 共演 し て い た 中井 貴一 が 「 女優 を やっ た 方 が いい 」 と 助言 し た そう 「 何 か を 手放し て やる 覚悟 も 必要 」 だ と 指摘 さ れ た という
Coverage	小池 栄子 が 、 本気 で 女優 の 道 に 出演 し て い た 大物 俳優 の 言葉 を 明かし た 「 そっち に 重き を 置い て 、 何 か を 手放し て やる 覚悟 も 必要 」 と アドバイス 「 女優 を やり たい なら 、 そっち に 重き を 置い て 、 何 も 教え て くれ なかっ た 」 と 語っ た
Parallel_Train	10 日 の 番組 で 小池 栄子 が 、 本気 で 女優 の 道 に 進む きっかけ と なっ た 大物 俳優 の 言葉 を 明かし た 新しく 女優 を やり たい なら 、 そっち に 重き を 置い て 、 何 か を 手放し て やる 覚悟 も 必要 と の こと 「 そこ を 言い訳 に すん な 』 と お 叱り だっ た と も
Sequence_Train	10 日 の 番組 で 小池 栄子 が 、 本気 で 女優 の 道 に 進む きっかけ と なっ た 大物 俳優 の 言葉 を 明かし た 「 バラエティー 出身 です し 」 と 逃げる 自分 に 、 「 女優 を やっ た 方 が いい 」 と アドバイス 「 何 か を 手放し て やる 覚悟 も 必要 」 だ と アドバイス し て いる

表 8.3　出力例 2（直列）.

記事 [5]	最近、65 歳以上のドライバーいわゆる“高齢運転者”への注目が高まっています。 警視庁が公開する“高齢運転者が関与した交通事故発生状況（平成 26 年中）”によると、交通事故の総件数は減少する一方で、高齢者が関与する割合は年々増加しています。 免許証の自主返納を促す流れが強まっていますが、都内ならまだしも、買い物や通院でクルマが生活に欠かせないような地域では難しいのが現状でしょう。 ただ、現在はクルマの安全技術も日々進化しています。なかでも緊急時の自動ブレーキは、スバルが先日発表した調査によると、搭載車では非搭載車にくらべて約 6 割の低減効果があるとのこと。いまでは軽自動車にも普及している技術ですが、危険を検知する方式に違いがあるのをご存じでしょうか？ 以前は赤外線レーダーのみで障害物を検知するに留まりましたが、現在はカメラと組み合わせた二段構えで車両だけでなく歩行者も検知するのが最新式。今回はそんな最新の安全装備を搭載した軽自動車をご紹介いたします。 スズキが採用する安全システムが「デュアルカメラブレーキサポート」です。 フロントウィンドウ上部にカメラを 2 つ用意。人の目と同じように左右それぞれの映像から対象のカタチと距離を捉え、そのサイズや輪郭から歩行者やクルマを認識して万が一の際は自動ブレーキが掛かります。なお、作動域は約 5〜50km/h（対歩行者の場合は約 5〜30km/h）。 また、10km/h 以下での徐行中には約 4m 先の障害物を認識し、アクセルが強く踏み込まれた際は自動制御して急加速を予防します。そのほかにも、カメラ方式のおかげで車線を検知できるため、約 60〜100km/h で走行中に車線をはみ出しそうになるとブザーで警告する車線逸脱警報機能も搭載されています。 （一部抜粋）
正解要約[6]	最新 の 安全 装備 を 搭載 し た 軽自動車 を 紹介 し て いる スズキ の スペーシア は ミニバン 並み の 広 さ を 備え て いる の が 魅力 ムーヴ は 基本 性能 を 徹底的 に 上げ 、 走り は コンパクト カー に 迫る ほど だ という
Coverage	高齢 運転 者 が 関与 し た 交通 事故 の 総 件数 が 関与 する 割合 へ の 注目 が 高まっ て いる 緊急 時 の 自動 ブレーキ が 発表 し た 調査 に よる と 、 高齢 運転 者 が 減少 する 搭載 に 欠か せ ない よう な 地域 で は 難しい の が 現状 だ という
Parallel_Train	高齢 運転 者 が 関与 し た 「 高齢 運転 者 」 へ の 注目 が 高まっ て いる 高齢 者 が 関与 する 割合 は 年々 増加 し 、 ドライバー の 注意 力不足 が 挙げ られる 搭載 車 で は 非 搭載 車 に くらべ て 約 6 割 の 低減 効果 が ある という
Sequence_Train	高齢 運転 者 が 関与 し た 安全 装備 を 搭載 し た 軽自動車 を 紹介 し て いる サイズ や 輪郭 から 歩行 者 や クルマ を 認識 し て 検知 する に は 「 デュアルカメラブレーキサポート 」 と の こと 映像 から 、 免許 証 の 自主 返納 を 促す 流れ が 強まっ て いる と の こと

表 8.4　出力例 3.

記事[7]	テレビアニメ「ドラゴンボール」シリーズの放送 30 周年を記念し、シリーズの歴代主題歌を網羅するベストアルバムの発売が決定した。主人公・孫悟空が必殺技・かめはめ波を放つイラストと、30 周年ロゴが組み合わされたアニバーサリー仕様のジャケットイラストが公開されている。 「アニメ『ドラゴンボール』放送 30 周年記念　ドラゴンボール 神 BEST」と題されたベストアルバムは、1986 年の初代オープニング主題歌「魔訶不思議アドベンチャー！」から劇場版、新作テレビシリーズ「ドラゴンボール超」の主題歌までを、CD2 枚組で収録する。2 月 24 日発売予定で、「ドラゴンボール」「ドラゴンボール Z」「ドラゴンボール超」など、全 7 作品のオープニング映像（「ドラゴンボール GT」以外はノンテロップ版）を収録した DVD が付属する初回限定盤が 3780 円、通常盤が 3000 円（いずれも税込み）。収録曲などの詳細は、日本コロムビアの特設サイトで発表予定。 購入者にはメーカー特典として B2 告知ポスターが配布されるほか、TSUTAYA では先着特典として、ジャケット絵柄ステッカーと、ジャケットイラストポートレートが抽選で 5 人に当たるキャンペーン応募はがきが配布される。 また、TSUTAYA では 3 月 12 日から「神 BEST」のレンタルもスタート。同アルバムを含む日本コロムビアがリリースする「ドラゴンボール」関連 CD をレンタルして応募すると、抽選で「神 BEST」ジャケットイラスト A4 ポートレイトが 5 人に当たるキャンペーンも行われる。
正解要約[8]	アニメ 「 ドラゴン ボール 」 の 30 周年 を 記念 し て 、 ベスト アルバム が 24 日 に 発売 歴代 の テレビ シリーズ や 劇場 版 の 主題歌 を 、 CD 2 枚 組 で 収録 し て いる 購入 者 に B 2 告知 ポスター が 配布 さ れ 、 TSUTAYA で は キャンペーン が 始まる
Coverage	「 ドラゴン ボール 」 シリーズ の 歴代 主題歌 を 網羅 アルバム の 発売 が 決定 した かめ はめ 波 を 放つ イラスト と 、 30 周年 ロゴ が 組み合わさ れ た アニバーサリー 仕様 ジャケット 絵柄 ステッカー と 、 ジャケットイラストポートレート が 抽選 で 5 人 に当たる キャンペーン 応募 はがき が 配布 さ れる
Parallel_Train	ドラゴン ボール が 歴代 主題歌 を 網羅 する ベスト アルバム の 発売 が 決定 し た 「 ドラゴン ボール 超 」 の 主題歌 まで を 、 CD 2 枚 組 で 収録 し た DVD が 付属 購入 者 に は メーカー 特典 として B 2 告知 ポスター が 配布 さ れる
Sequence_Train	アニメ 「 ドラゴン ボール 」 シリーズ の 歴代 主題歌 を 網羅 する ベスト アルバム の 発売 が 決定 購入 者 に は 先着 絵柄 ステッカー と 、 ジャケットイラストポートレート が 抽選 で 5 人 に当たる キャンペーン 応募 はがき が 配布 さ れる TSUTAYA で は 3 月 12 日 から 「 神 BEST 」 の レンタル も スタート する

表 8.5　出力例 4.

第 9 章　おわりに

本研究では Livedoor News からデータを収集し，3 行要約に注目した大規模なニュース文書要約データセットの構築を行った．さらに，3 行要約の文書構造に着目したアノテーションを行なった．また，アノテーションで付与されたタグの要約構造に適応したモデルの検討を行った．最後に，3 行要約に着目した評価指標の提案をした．

本研究では，fine-tuning に用いる分類器と $Merge$ モデルを構築する際に用いた分類器のどちらも，簡単な分類器であったが，記事にはカテゴリ等の情報も付与されているので，それらを考慮する分類モデルが必要である．また，要約モデルも 3 文目のスコアが低い結果となっているため，モデルのネットワークに要約構造を考慮するような構造が必要である．

謝辞

本論文の執筆において，指導教員の小町先生にはとてもお世話になりました，感謝いたします．

また，本論文の副査を引き受けてくださった石川博教授，片山薫准教授に感謝いたします．

今回の論文を書くにあたり，研究室の同期・先輩・後輩には研究のアドバイスをもらい助けていただきました．ありがとうございました．

参考文献

[1] R. Nallapati, F. Zhai, and B. Zhou, “Summarunner: A recurrent neural network based sequence model for extractive summarization of documents,” Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp.3075–3081, 2017.

[2] C. Li, X. Qian, and Y. Liu, “Using supervised bigram-based ILP for extractive summarization,” Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp.1004–1013, 2013.

[3] W.C. Mann and S.A. Thompson, “Rhetorical structure theory: Toward a functional theory of text organization,” Text-Interdisciplinary Journal for the Study of Discourse, vol.8, no.3, pp.243—281, 1988.

[4] A.M. Rush, S. Chopra, and J. Weston, “A neural attention model for abstractive sentence summarization,” Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.379–389, 2015.

[5] I. Sutskever, O. Vinyals, and Q.V. Le, “Sequence to sequence learning with neural networks,” Advances in Neural Information Processing Systems 27, pp.3104–3112, 2014.

[6] R. Nallapati, B. Xiang, and B. Zhou, “Abstractive text summarization using sequence-to-sequence RNNs and beyond,” Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, pp.280–290, 2016.

[7] A. See, P.J. Liu, and C.D. Manning, “Get to the point: Summarization with pointer-generator networks,” Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp.1073–1083, 2017.

[8] C.-Y. Lin, “ROUGE: A package for automatic evaluation of summaries,” Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, pp.74–81, 2004.

[9] D. Gillick and B. Favre, “A scalable global model for summarization,” Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing, pp.10–18, 2009.

[10] T. Hirao, Y. Yoshida, M. Nishino, N. Yasuda, and M. Nagata, “Single-document summarization as a tree knapsack problem,” Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp.1515–1520, 2013.

[11] 駿　田中，遼平笹野，大也高村，学　奥村，“要約長，文長，文数制約付きニュース記事要約，” 言語処理学会第 22 回年次大会，pp.341–345，2016.

[12] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, “On using very large target vocab-

ulary for neural machine translation," Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp.1–10, 2015.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol.9, no.8, pp.1735–1780, 1997.

[14] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," Proceedings of International Conference on Learning Representations, 2015.

[15] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer networks," Advances in Neural Information Processing Systems 28, pp.2692-2700, 2015.

[16] Z. Tu, Z. Lu, Y. Liu, X. Liu, and H. Li, "Modeling coverage for neural machine translation," Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp.76–85, 2016.

[17] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," J. Mach. Learn. Res., pp.2121–2159, 2011.

[18] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," Proceedings of the 30th International Conference on Machine Learning, vol.28, no.3, pp.1310–1318, 2013.

発表論文リスト

査読あり会議

[i] T. Kodaira, T. Kajiwara and M. Komachi, "Controlled and balanced dataset for japanese lexical simplification," Proceedings of the Association for Computational Linguistics 2016 Student Research Workshop, pp1–7, 2016.

査読なし会議

[ii] 小平 知範, 梶原 智之 and 小町 守, "均衡コーパスを用いた日本語語彙平易化データセットの構築," 言語処理学会第 22 回年次大会, pp.258-261, 2016.

[iii] 小平 知範 and 小町 守, "TL;DR 3 行要約に着目したニューラル文書要約," WebDB Forum, 2017.

付録 A　各タイプの記事例

「ビッグマック指数」をご存じですか？
各国のマクドナルドで販売されているビックマックの単価は、材料費や店の光熱費、店員の給料といった各国の経済事情が集約して決定されているのです。その単価を比較し指数化することで、それぞれの国の経済力が推し量れるのです。
これはマクドナルドが同じ商品を全世界に販売しているからこそ出てきた考え方です。実はこの考え方、なんとデザインの分野にも転用ができるのです。それが「マクドナルドＨＰデザイン指標」です。
マクドナルドは全世界で統一したブランドイメージの基にほぼ同等の商品を販売しています。それにも関わらず各国のホームページのデザインは異なっているのです。この同一情報に対する表現の違いを比較することで、それぞれの国のベーシックなデザインの思想を推し量ることができるのです。
■ドイツ（サムネイル左上）
たくさんの情報がバランスよく整理整頓されており、見やすさと美しさの両立が感じとれます。
■フランス（サムネイル右上）
細い書体で大きくといった UI グラフィクスのトレンドを取り入れており、モダンさが感じとれます。
■イタリア（サムネイル左下）
豊かな色合いの写真とにぎやかな商品ロゴで、大人の楽しさが感じとれます。
■アメリカ（サムネイル右下）
黒バックに商品を載せて太い書体を重ねることで、強く明快な感じを受けます。スマートフォンＯＳのトレンドを取り入れており、ＯＳ大国の所以を感じとれます。
■日本
傾向はアメリカに近いですが、加えて情報をたくさん載せなければという意図が垣間見えます。
このように各国のホームページごとにデザインに特徴があり、そこからそれぞれの国のベーシックなデザインの思想が垣間見えます。この「マクドナルドＨＰデザイン指標」を数年ほど観察し、変遷を見える化することで、もしかしたらデザインのトレンドが細かくどこからどこへ流れていくのかが分かるかもしれませんね。
（Betonacox Design）

表 A.1　並列タイプの記事例.

円高への警戒感、金利の低下、英国の EU 離脱、中国経済の失速不安などで、秋口に入るまで株式市場は停滞が続いていたが、その中でひとり気を吐く株がある。いったいどんな産業だろうか。

「米ゴールドマン・サックスは 2025 年に VR・AR 関連の世界市場が 950 億ドル（約 9 兆 5000 億円）に拡大すると予想する」（日本経済新聞 2016 年 9 月 29 日）。記事は任天堂、ソニー、バンダイナムコホールディングスといったゲーム関連株の堅調ぶりを伝えている。今、投資家が注目しているホットなテーマ。それが、ゲーム業界における「仮想現実（バーチャルリアリティ・VR）」だ。

10 月 13 日、ソニー・インタラクティブエンタテインメント（SIE）が発売したゴーグル型端末機器「以下 PS VR）」は発売と同時に売り切れ、話題沸騰中。発売の前にも 3 回の予約が設けられたが、全て「瞬殺」開始と同時にほぼ完売。生産が追いつかない人気ぶりを見せている。

久々の大型ヒットの予感に業界もお祭り気分だが、よく考えてみよう。「10 兆円市場」といえばコンビニ全体の売上高と同規模だ。そんな巨大な市場が本当にゲーム産業から生まれようとしているのだろうか？　もしそうだとすれば、「ゲーム」の枠に留まらない何か凄い起爆力が潜んでいる？　取材はまず、「PS VR」を実体験することからスタートした。

◆全視野を埋め尽くす仮想世界

助手席に座ると高速で走り始める車。疾走する車のドアを左手で押し開けてみる。路上の白線がビュンビュンと凄いスピードで後方へ飛び去っていく。後ろから追いかけてくる車が見えた。背筋がゾクッとする。

怪しげなバイクと車が近づいた、と思ったらいきなり銃を乱射してきた。慌てて応戦。銃を握り、弾を装填して相手のタイヤを狙って撃ち返す……従来のシューティングゲームとケタ違いに生々しいのは、360 度、全方向を映像に取り囲まれているから。私が頭を動かすと、景色はめくるめくほど変わっていく。まさに現場のリアル感。

「その実在感こそ、PS VR の魅力であり最大の特徴です」と SIE グローバル商品企画部・高橋泰生氏（41）は言う。これまではいくら巨大画面であってもフレームに囲まれていて、それが現実と非現実の「境」になっていた。しかし PS VR は視野全てをゲーム世界が埋め尽くす。

「広い視野角が臨場感を高めているだけでなく、VR ヘッドセットには頭の動きや傾きを外部の専用カメラが検知する LED ライトを搭載。さらに、ジャイロセンサーや加速度センサーを組み合わせることで頭がどう傾いたか、視線をどこへ投げたか、向きや位置を正確に測定し、映像に反映させていきます」

開発当初は画面に液晶を使用。しかし最終的に有機 EL を採用しました、と高橋氏は続ける。

「VR の特性に最適な素材をと考え、オリジナルパネルを作り上げたのです。有機 EL を採用することで画面の応答速度を 0.018 秒未満にすることが可能になりました。ブレや残像感のない、キレのある表現が実現できました」

VR による実在感や臨場感を創り出す上で、最も大事になることとは何でしょう？
「視覚情報と体の情報とをマッチングさせることですね。目というのは実はものすごく繊細な器官。たった 0.02 秒ほど頭の動きに映像が遅れただけでも違和感につながってしまう。そうした視覚と体感のズレが、『VR 酔い』を生んでしまうのです」
「VR 酔い」とは、初めて聞く言葉だ。
「車酔いのような違和感や不快感をいかに少なくできるか。技術陣が最も注力した点なんです」と高橋氏は言う。
「頭に被る VR ヘッドセットも例外ではありません。敢えて後方に重りを入れて、頭の真上に重心がくるよう設計しました。装着感や違和感を最大限減らすための工夫です。ソフトも体の動きと映像との間に違和が生じないように、コンテンツ制作サイドと綿密に詰めていきました」
◆「399 ドルターゲット」の狙い
では、映像以外ではどうか。SIE ならでは、の技術はどこに活かされているのか？
「360 度、あらゆる方向から立体的に音が聞こえてくる 3D オーディオを独自に開発。実在感を際立たせる効果があります」
もう一つ、手の動きを伝えるコントローラーも重要だ。私がコントローラーを動かすと、映像の中の手が滑らかに動いていく。触る、握る、ドアを開ける、銃を撃つ、といった動作が映像内で次々に実現されていく。視覚、聴覚、触覚。「仮想の夢」から「覚めない」よう、とことん作り込まれた、全く新しい娯楽世界がそこにあった。
と、ハイスペックで綿密に作り込まれた新製品だが、意外だったのはその値段。何と 4 万 9980 円（専用カメラ含む。PS4 は別途）、競合他社の VR と比べても破格の安さだ。振り返れば 1980 年代、「VR」の言葉が登場した頃のヘッドマウントディスプレイは 300 万円もしたのに……。
「PS VR のプロジェクトが始まったのは 2012 年。その後『399 ドル』ターゲットを設定しました。普及させるために PS4 の本体価格（2 万 9980 円〜3 万 4980 円）を大きく超えないことを意識してきました」
実は低価格化の背後に社会の劇的な変化もあった。スマホの普及によって傾きや動きを感知するセンサー、ディスプレイ等電子部品の価格が下がり、手軽な価格での VR が実現可能になったという。
「いつかは一家に一台、テレビのように PS VR を行きわたらせたいと真剣に考えています。なぜならゲームに限らず、VR が表現する実在感こそコミュニケーションツールの土台になる、と思っているからです」
◆ VR 文化を創り出せるか
例えば北極圏でオーロラを体感するコンテンツ。遠出が難しくなった高齢者も、海外旅行のリアルな感動を VR によって楽しめる日が来た。
「そのためにはゲームをしない人も違和感無く、ごく自然に使えるツールであることが何よりも大切になるのです」

あたかもそこにいるような実在感を伝える VR 技術。その可能性は、とてつもなく広い。「臨場感」は人間のコミュニケーションの全てに関わってくるからだ。遠隔地の人と顔をつきあわせたような会話も、新しい家の間取りを体感することも VR で可能になる。医療なら手術のシミュレーション、企業では危険な作業や特殊な場所を想定した操作訓練……応用範囲は限りない。

かつて、「音楽を持ち歩く」というまったく新しい「体験文化」をウォークマンによって創り出したソニー。今後社会で活用の裾野が大きく広がりそうな VR 技術を、「VR 文化」へと、どのように仕上げていくのか。全ては PS VR の展開にかかっている。

表 A.2: 直列タイプの要約例.

今年もボジョレー・ヌーヴォー解禁！　まあ、昔ほど盛り上がってはいないのだが、それでも出来栄えが気になるところ。今年は、果実味を楽しめる仕上がりになっているそうだ。
そこで、大手コンビニ（セブンイレブン・ローソン・ファミリーマート）各社が販売している、ボジョレー・ヌーヴォーを飲み比べてみた。って、あれ？　作り手は全部一緒。味に違いなんかあるのかな？
・各社のボジョレー
飲み比べたのは、次の 3 種。セブン：ジョルジュ デュブッフ「ボージョレ・ヴィラージュ ヌーヴォー＜クー・ド・プレス＞ 2016」2480 円
ローソン：ジョルジュ デュブッフ「ボジョレー・ヴィラージュ ヌーヴォー セレクション プリュス」3000 円
ファミマ：ジョルジュ デュブッフ「ボジョレー・ヴィラージュ ヌーヴォー 2016」2480 円
・見た目に違いはない
ボトルを見たところ、特に大きな違いを感じない。プラスチックコップに注いでみたところ……。色目を見ても全然違いを感じない。ついでに香りもほぼ一緒。味にも違いはないのだろうか？
当編集部 7 人で飲み比べをしてみたところ……。やっぱり予想通りの結果に。
・編集部メンバーが 1 番美味しいと感じたボジョレー・ヌーヴォー
セブン：サンジュン・和才
ローソン：佐藤・羽鳥
ファミマ：りょう、せいじ・Yoshio
意見がバラバラになったうえに、全員が自分が 1 番美味しいと感じた理由を「飲みやすい」と評した。これはもう微々たる差でしかないように思うのだが。あとは好みの問題。多少渋味が強いもの（セブン）が好みか、多少後味がスッキリしたもの（ローソン）が好みか、それとも多少酸味が強いもの（ファミマ）が好みかの違いしかない。
・結構売れているらしい
それでもローソンのボジョレーは人気が高く、予約分完売で店舗在庫を確保できなかったお店もあるそうだ。ボジョレー熱は以前ほどアツくないものの、それでも売れ行きは良いようである。興味のある人は、それぞれの商品を購入して、飲み比べてみてはいかがだろうか。
Report：佐藤英典
Photo：Rocketnews24

表 A.3　列挙型並列の記事例.

TBS「ジョブチューン ～アノ職業のヒミツぶっちゃけます！」（23 日放送分）では「引退した今だからぶっちゃけます！スペシャル」と題し、昨年引退した元中日ドラゴンズ・山本昌氏がゲスト出演。現役時代、星野仙一氏に殴られたエピソードを明かし、周囲を驚かせた。
「記憶に残るケガ」を訊かれた山本氏は、2006 年 10 月の日本シリーズで骨折しながらも登板したエピソードを披露。「痛いことは痛いけど投げれちゃったもんだから」と振り返る。
すると山本氏は、1987 年～1991 年の星野仙一監督時代について「投げ終わってベンチ裏に来いと言われて多少かわいがられまして」と暴露。「顔が腫れ過ぎて降板した」と驚きの事実を伝えた。
「だいぶ殴られた」という山本氏は「5 発目までは自分で数えてましたけど、その後、覚えてない。途中から温かくなるだけ」と苦笑い。また殴られた理由については「井上（真二）選手に 2 本ホームランを打たれた。そしたら“お前、名前も知らん選手にホームラン打たれやがって”って」と説明したが、山本氏は「その年（井上は）オールスター行ってるし」と心の中で反論していたという。
しかし、「厳しいだけじゃない。叱られれば次絶対使ってくれた」と星野氏のフォローを忘れなかった山本氏。「いつも一番感謝していますし、引退を決めた時も一番最初に電話したのは星野監督」と語った。

表 A.4　文分割型直列の記事例.