

# 575の音韻的読みやすさを付与した学術論文の要約文自動生成手法

安部 文紀<sup>†</sup> 寺田 実<sup>†</sup>

<sup>†</sup> 電気通信大学情報理工学研究所 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: †{abe-fuminori,terada.minoru}@uec.ac.jp

あらまし 学術論文の投稿と検索に特化した Web サービスの台頭によって論文を大量に入手できるようになった。これに伴い、論文を読むべきかどうか判断する機会と概要を思い出すために論文を読み返す機会が増えた。一般的にこのような機会にはタイトルやアブストラクトが用いられるが、タイトルとアブストラクトよりも短く、かつ印象に残り易い文章が論文に付与されていれば、論文を読むべきか判断する手間と概要を思い出す手間が軽減することが期待される。そこで本研究では、古来より親しまれてきた 575 の持つ音韻的読みやすさを付与した要約文が望ましいと考え、学術論文の新たな要約手段として 575 形式のキャッチフレーズで論文を表現する 575 自動生成手法を提案する。評価実験では提案手法により自動生成した 575 (論文 575) とタイトルを比較するユーザスタディを行った結果、タイトルよりも少ない文章量で論文概要を表現できることを確認した。また、論文 575 から受ける印象を人生成の 575 と比較したところ、手製のものより劣る結果となった。印象に残り易い論文 575 を生成する手法の探求を今後の課題とする。

キーワード 文書要約, 自動要約, 575 生成, 特徴語抽出

## 1. はじめに

研究をするにあたって学術論文を調査することは必要不可欠である。近年は Google Scholar<sup>(注1)</sup> のような論文検索エンジンや, arXiv<sup>(注2)</sup> のような論文投稿 Web サイトの台頭によって所望の論文を大量に、かつ容易に入手できるようになった。これに伴い、論文を読むべきかどうか判断する機会と概要を思い出すために論文を読み返す機会が増えた。このような機会では一般的にタイトルやアブストラクトが用いられるが、タイトルとアブストラクトよりも短く、かつ印象に残り易い文章が論文に付与されていれば、論文を読むべきか判断する手間と概要を思い出す手間が軽減することが期待される。

論文をタイトルとアブストラクトよりも少ない文章量で、かつ印象に残り易く表現するために、本研究では古来より親しまれてきた俳句や川柳といった 575 の音韻的読みやすさを付与した要約文自動生成手法を提案する。例えば本稿を提案手法に入力した場合、図 1 のような要約文が出力されることを想定する。



図 1 本論文を提案手法 (論文要約 575) に入力した場合に想定される 575 の出力例

本研究は論文を少ない文章量で、かつ印象に残り易く要約することで論文読者に貢献する。そして、論文に付与するキャッチフレーズやキーワードの生成を代行、考案を支援することで論

文著者に貢献する。科学技術論文では、研究のダイジェストとなる画像やイメージ図、もしくは研究のキーワードが論文の冒頭に付与される習慣があるため、将来的にはそのようなダイジェストにおいて「論文要約 575」が併用されることを目指す。

## 2. 関連研究

### 2.1 575

本研究では要約文の生成に 575 の形式を用いるが、575 には二つのメリットが存在する。ひとつは「音韻的に読みやすいこと」、もうひとつは「記憶に残りやすいこと」である。

まず二つ目のメリットについて述べる。575 は作成される際、五音節と七音節を基調とされており、言い換えると基準となる音数の規則が存在する。これを五七調・七五調と呼ぶ。坂野 [1] によると、五音と七音が心地良いと結論づけており、理由は、二音一単位の「律拍」という概念に基づいて二音、四音、八音、... (2 の冪乗) は繰り返すとリズムカルでありスムーズであるためと述べている。575 は上五、中七、下五のそれぞれが 2 の冪乗の音数にはなっていないが、五音の後に三音の休止 (空白) が入り、七音の後には一音の休止が入り、それぞれが八音となっている。575 は律拍のなかでも八音を基本の単位としていて八音に満たない部分については休止が入るようになっていて述べている。また、渡部ら [2] は、古来より五七調・七五調のリズムが良いとされていることについて疑問を呈し、現代人でも五七調・七五調がリズムカルに感じるかどうかを実証実験した。結果は七五調、五七調、六六調、四八調、八四調の順に心地良いと被験者が評価したため、「五音七音がリズムカルと感じるのは、単になじみのある型の持つ心地よさというだけではなく、五音七音という音の構成がリズム知覚に影響している」と結論づけている。

次に二つ目のメリットについて述べる。越場ら [3] は、リズムカルな五七調・七五調は記憶に残りやすいという仮説を掲げ、五

(注1): <https://scholar.google.co.jp/>

(注2): <https://arxiv.org/>

七調が暗記学習に効果があるかを実証実験した。歴史の暗記問題を韻律を考慮しない語呂合わせで覚えてもらった場合と、韻律を考慮した五七調の語呂合わせで覚えてもらった場合の正解数を比較したところ、後者の正解数の方が多かったため、「五七調は暗記学習に効果があることがわかった」と結論づけている。

まとめると、575 の五七調はリズムカルで耳に残りやすい特徴があるため、文章が記憶に残りやすいメリットがあると言える。学術論文は長文で堅苦しい言い回しがあるため、論文を身近な五七調で表現することによって読みやすく、印象に残りやすくなると考える。

ここで、575 の形式が読みやすく、印象に残りやすいことを利用して注意喚起や啓発イベントにて標語を広告する際に五七調が採用されている例をいくつか紹介する。北海道旭川市で行われた選挙の際には、投票へ行くことを促すためのスローガンを募集し、採用されたスローガンのほとんどが五七調であった<sup>(注3)</sup>。また、広島県広島市では毎年人権啓発イベントを行っており、その際にスローガンを募集している。採用されているほとんどが五七調である<sup>(注4)</sup>。その他、交通安全の標語でも五七調は多く採用されている<sup>(注5)(注6)</sup>。五七調が知覚に訴えるリズムの良さと五七調の持つ印象に残りやすさは、周知の事実であることが実績を通して確認できる。

## 2.2 学術論文から要約文、タイトルを生成する研究

学術論文を短く表現するために要約文やタイトルを生成する研究が取り組まれてきた。要約文生成の研究には「抽出型要約」と「生成型要約」が存在する。抽出型要約では論文中の各文を重要度の高い順に並び替え、いくつか抽出したものを要約文としている。生成型要約では論文から抽出した文を要約文として扱うことはせず、人間が要約を作る際に行う言い換え表現のような単語列の再構成を機械学習によって行い、生成した新しい単語列を要約文としている。抽出型要約の重要度の導出において、論文中の文や句間の関係を表す談話構造に着目し、各文から抽出した談話構造に基づく特徴量を用いてアブストラクトとの類似度を測る手法[4]を用いた研究がある。他にも、Page-Rank アルゴリズムから着想を得た LexRank を用いて、入力する論文の主張に関連する語句部分を、その論文を引用している論文から抽出する手法[5]も存在する。生成型要約の単語列の生成において、機械翻訳のタスクで用いられるニューラルネットワークの Encoder-Decoder モデルを論文要約のタスクに適用し、全文を一度に処理するのではなく論文の章構造に基づきセクションごとに処理する方法を提案した研究[6]が存在する。要約文生成の研究はほとんどが Extended Abstract のような長文を生成する研究であり、論文内容の理解支援を目的としている。本研究は、論文内容の理解支援ではなく、内容理解の前段階で行う論文を読むべきかどうか判断することの支援を目的とする。

(注3): <http://www.city.asahikawa.hokkaido.jp/kurashi/461/463/d053708.html>

(注4): <http://www.city.hiroshima.lg.jp/www/contents/1268627303816/index.html>

(注5): <http://www2.tadsa.or.jp/anzen2017s/>

(注6): <http://www.mainichi.co.jp/event/aw/anzen/slogan/archive.html>

タイトル生成の研究は一種の要約文生成とみなして取り組まれてきたため、要約文生成の研究と同様に本文から抽出した最重要な一文をタイトルとする抽出型の手法[7]や、Encoder-Decoder モデルを用いてアブストラクトから新しい単語列を生成する生成型の手法[8]が存在する。タイトルは要約文と比較すると文章量は少ないが、本研究は五七調で構成される17音で要約するため、575と比較すると一覽性に欠ける。また、2.1で述べた通り575の五七調は記憶に残りやすい利点を持つため、印象に残りやすさの点においてもタイトルよりも優位性がある。

## 2.3 575 生成の研究

自然言語処理技術を用いて機械的に575を生成しようと試みた研究がいくつか存在する。RafalらはWebに存在する文書を言語資源として使い、Web検索ヒット数とユーザが選択するブログ記事における単語の出現頻度を基に俳句に用いる単語を決定し、オリジナルの文法テンプレートに抽出した単語を割り当てて俳句を自動生成する手法[9]を提案した。DanielらはWikipediaを言語資源として使い、単語 N-gram モデルによって俳句を自動生成する手法[10]を提案した。Yealらは俳句協会のWebサイトや童話、松尾芭蕉の俳句等を言語資源として使い、連想語ネットワークを利用して俳句に用いる単語を決定し、品詞 N-gram モデルによって俳句を生成する手法[11]を提案した。Martinらは、俳句に多用される550単語を格納したシード単語辞書を用意し、言語資源であるブログ記事から俳句に使う文字列を検索するためのクエリとして用いた。そして、ヒットした単語を含む文をブログ記事から抽出し、Vector Space Model と TF-IDF (Term Frequency - Inverse Document Frequency) を用いて俳句に用いる文を選択、結合する手法[12]を提案した。

575生成の研究は、人手で作ったものにどれだけ近付けられるかを目的に、そして俳句の美しさをどれだけ持てるかを課題に行われてきた。そのため、575の自動生成を要約文の生成と見なして取り組んだ研究が存在しない。また、本研究と同様に入力文書の存在を仮定している575生成の研究も存在するが、生成した575の候補から最終的なひとつに絞り込む手法の提案はされていない。本研究では入力論文から要約文とみなした575の候補を生成し、最終的なひとつを決定するまでの一貫した575自動生成手法を提案する。そして、「少ない文章量で論文を表現できるか」、「印象に残り易いか」を評価実験で確認する。

## 3. 予備実験: 学術論文を575で表現する実験

### 3.1 概要

自動生成した575を用いて「少ない文章量で論文を表現できるか」、「印象に残り易いか」を評価実験で確かめる前に、まずは学術論文の要約文を575形式のキャッチフレーズで表現することの有効性を確認する必要があると考えた。そこで、ソフトウェア工学関連の学会論文集一冊から筆者が人手で作成した575形式の要約文を評価者に読んでもらう実験を2つ行った。具体的には、学術論文から作成した575が、論文にあらかじめ付与されているタイトルと同じくらい、もしくはそれ以上に論文を表現できるかをユーザスタディにより確認した。

### 3.2 準備

実験に用いる学術論文とその 575 を用意するために、WISS 2016<sup>(注7)</sup> 論文集<sup>(注8)</sup> に掲載の登壇発表とデモ・ポスター発表の論文 (合計 82 本) を読み、筆者の主観で各論文について 575 形式のキャッチフレーズ (全 82 個) を作成した。論文タイトルと作成した 575 の例を以下に示す。

- 3D プリンタのための Paper User Interface<sup>(注9)</sup>  
オブジェクト スキャンするだけ 二面図を
- 着ぐるみ非装着時の着ぐるみポーズ練習システム<sup>(注10)</sup>  
着ぐるみの ポーズ練習 支援する
- 物理ベースサウンドデザインのための例示ベースインタフェース<sup>(注11)</sup>  
音分布 例示ベースで 生成し

### 3.3 方法

被験者 5 人に「実験 1: 学術論文のアブストラクトとその 575 を結びつける実験」、「実験 2: 学術論文のアブストラクトとそのタイトルを結びつける実験」を行ってもらった。各実験、ひとつのアブストラクトにつき正解 (実験 1 では正解 575、実験 2 では正解タイトル) を 1 つ、不正解を 4 つ含んだ解答候補を 5 つ提示し、5 問 (5 本の論文) について回答してもらった。実験で扱うアブストラクトは WISS 2016 論文集に掲載されている 82 本の論文からランダムに抽出した 10 本の論文とし、5 本を実験 1 に、残り 5 本を実験 2 に用いた。

学術論文のタイトルは、一般に論文内容の主張点を的確に表したもとのとして扱われるため、575 のアブストラクト正解率 (実験 1) がタイトルのアブストラクト正解率 (実験 2) と同等、もしくはそれ以上であれば学術論文の概要を 575 (タイトルより少ない文章量) で表現することの有効性を示せることとした。

### 3.4 結果と考察

実験 1 と実験 2 の正解率が両方とも 100 % となった。したがって、学術論文を 575 で表現することは有効であり、また、論文のタイトルと同等の役割 (論文内容の主張点を的確に表す) を持つことがわかった。まとめると、575 形式のキャッチフレーズで学術論文の概要を表現することは可能である。

本予備実験の準備作業として行った、論文から 575 を人手で作る工程で得た知見を、提案手法により生成する 575 の候補から適切なひとつを出力する際の評価指標に適用することとした。代表的な知見を以下に示す。

- タイトルに出現する単語は 575 に採用しやすい
- アブストラクトに出現する単語は 575 に採用しやすい
- 係り受け関係が適切な 575 は論文要約 575 として採用しやすい

## 4. 提案手法

### 4.1 概要

論文を 575 の形式で表現するにあたって、「内容的に正しい 575」と「面白い 575」の 2 つの方向性があると考えが、本研究では内容的に正しい 575、つまり学術論文を正確に表現することを重視して取り組む。

ここで、提案手法の要件を再掲する。

- 「少ない文章量で論文を表現できる」
- 「印象に残り易い」

1 つ目の「少ない文章量で論文を表現できる」について。提案手法は、音韻的に読みやすい形式で要約文を生成するため、手元にある論文を読むべきかどうかスクリーニングする際や、論文の概要を把握したい場面で役に立つことが期待される。そして 2 つ目の「印象に残り易い」について。馴染みのある形式 (五七調) で論文を表現することには、2.1 にて述べた 575 の持つメリットによって論文を印象づけて記憶に残す効果があるため、読み返したい論文がある場面や概要を想起したい場面でも役に立つことが期待される。

提案手法では学術論文 (和文) の PDF ファイルを一本入力すると、(1) PDF ファイルを全文テキスト化、(2) テキストの前処理を行わずに特徴語を抽出、(3) 特徴語から 575 形式のキャッチフレーズを生成してそれを出力する。本研究で提案する部分は (2)、(3) であり、具体的には 3 ステップを踏む (図 2)。

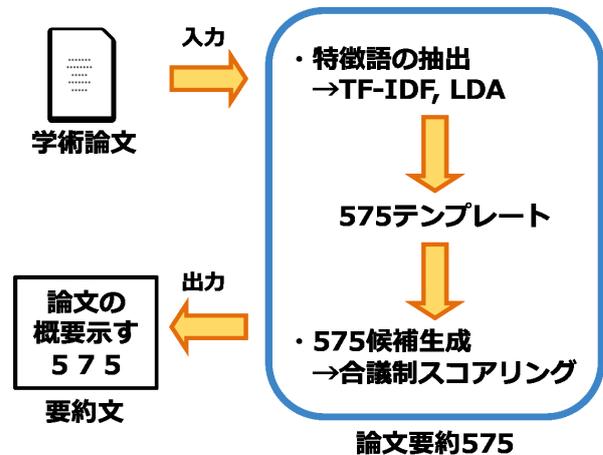


図 2 学術論文の 575 を生成するまでのステップ

- 特徴語の抽出
- 575 テンプレートに特徴語を割り当てて 575 の候補を生成
- 575 の候補を合議制スコアリングによって評価し、ランク付け

以下、各ステップについて説明する。

### 4.2 特徴語の抽出

提案手法 (論文要約 575) に入力された学術論文の特徴語の決定は、以下に示す特徴語抽出手法を用いて文中の名詞をスコアの低い順にランク付けすることで行う。なお、名詞に限定した理由は、575 テンプレート (後述する) に含まれる内容語がほ

(注7): <https://www.wiss.org/WISS2016/>

(注8): <http://www.wiss.org/WISS2016Proceedings/>

(注9): <http://www.wiss.org/WISS2016Proceedings/oral/25.pdf>

(注10): <http://www.wiss.org/WISS2016Proceedings/oral/09.pdf>

(注11): <http://www.wiss.org/WISS2016Proceedings/oral/10.pdf>

とんど名詞であったためである。本研究の特徴語抽出手法では、「文の表層情報に基づく手法」と「文の潜在情報に基づく手法」を組み合わせた手法を用いた。

文の表層情報に基づく手法では、単語の出現頻度に基づく単語重要度を用いた。特に TF-IDF (式 (3)) を用い、TF には入力する論文における単語出現頻度 (式 (1)) を、IDF には論文全文をピリオドで区切って抽出される文の集合を文書集合としたときの単語の文出現頻度 (式 (2)) を用いた。これは、予備実験にて人手で 575 を作成したところ、論文中の高頻度語が 575 に使われることが多かったためである。式 (1) の  $n_{w,d}$  は、入力論文  $d$  における単語  $w$  の出現回数、 $\sum_{s \in d} n_{s,d}$  は入力論文  $d$  の単語総数である。式 (2) の  $N_d$  は、入力論文  $d$  に存在する文の数、 $DF_{w,d}$  は論文  $d$  において単語  $w$  の存在する文の数である。TF-IDF の適用により、各単語の重要度がスコアとして得られ、スコアの高い単語ほど重要となる。

$$TF_{w,d} = \frac{n_{w,d}}{\sum_{s \in d} n_{s,d}} \quad (1)$$

$$IDF_{w,d} = \log \frac{N_d}{DF_{w,d}} \quad (2)$$

$$TF-IDF_{w,d} = TF_{w,d} \times IDF_{w,d} \quad (3)$$

文の潜在情報に基づく手法では、文書に潜在する意味に基づく単語重要度を用いた。特に、トピックモデリングの潜在的ディリクレ配分法 (Latent Dirichlet Allocation, LDA) [13] を採用し、入力論文の持つトピック (意味) を表現する名詞の重要度を求めるために用いた。LDA において、潜在トピックの推定手法には変分ベイズ推定 (Variational Bayesian Estimation, VBE) [14] を用い、トピック数は階層ディリクレ過程 (Hierarchical Dirichlet Process, HDP) [15] を用いることで無限化した。文書集合には、言語処理学会論文誌 L<sup>A</sup>T<sub>E</sub>X コーパス<sup>(注12)</sup> (565 本) を用い、これを入力する論文を加えたものを文書集合とした。

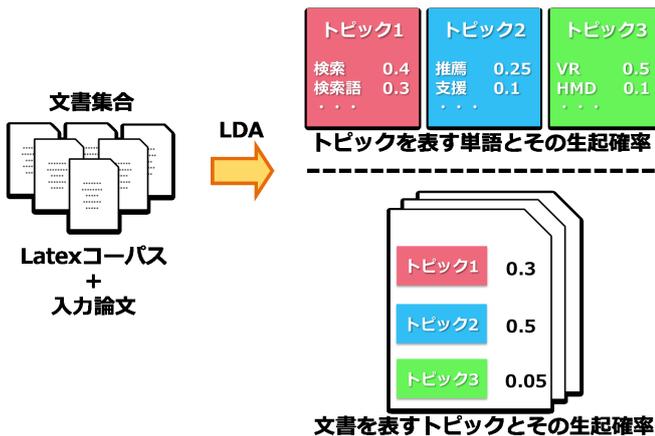


図 3 LDA によるトピック分布と単語分布の生成例

LDA では、1 つの文書に複数のトピックが存在し、1 つの単語が複数のトピックに存在しうる (図 3)。そのため、LDA の適用によって、

(1) 各潜在トピックを表す単語が生起確率とともに得られる (図 3 の「トピックを表す単語とその生起確率」)

(2) 各論文が持つ潜在トピックが生起確率とともに得られる (図 3 の「文書を表すトピックとその生起確率」)

本研究では、(2) において入力論文のトピックを占める割合が最も大きいトピックを入力論文の持つトピックとし、そのトピックにおける各単語の生起確率を重要度とした。

文の表層情報と潜在情報を組み合わせた手法には、unigram-rescaling [16] を適用した。unigram-rescaling では、式 (4) によって表層情報  $u$  に基づいた単語  $w$  の生起確率が、同時に潜在トピック  $z$  が与えられたときの生起確率に調整される。これによって、単語の出現頻度に入力論文の持つトピックも考慮したスコアが得られる。

$$p(w|u, z) \propto p(w|u) \left( \frac{p(w|z)}{p(w)} \right) \quad (4)$$

$u$  は、単語  $w$  より前に現れる単語であり、 $p(w|u)$  は何単語まで考慮するかによって変動する N-gram 確率である (一般に  $N \geq 2$ )。提案手法では、一般的な N-gram 確率 ( $N \geq 2$ ) を使用せずに TF-IDF のスコアをそのまま用いることとしたため、 $N = 1$  の unigram 確率 (語順を考慮しない単純な単語の出現頻度) にしたことと同義である。 $p(w)$  は単語  $w$  の uni-gram 確率であるため、入力する論文における出現頻度をそのまま適用した。 $z$  は潜在トピックであり、 $p(w|z)$  はトピックにおける単語  $w$  の生起確率である。提案手法では、LDA によって得られる入力文書における最大占有率のトピックの単語分布を適用した。

はパラメータである。

ここで、パラメータ の重み決定アルゴリズムを述べる。の重みは、以下の手続きによって最適値を決定した。

(1) WISS2016 の学術論文からランダムに 5 本の論文を選出し、各論文について (2) ~ (4) を繰り返す

(2) 1 本の論文について、 の値を -1.0 から 1.0 の範囲で 0.1 刻みで変化させながら (3) を繰り返す

(3) unigram-rescaling によって導出した特徴語の上位 20 個が論文のタイトルとアブストラクトに存在する数 (最大 40 個) をポイントとする

(4) ポイントの大きい順に に順位をつける (最高 1 位、最低 20 位)

(5) それぞれの順位になった回数を用いて  $nDCG_{20}$  (Normalized Discounted Cumulative Gain) [17] で の各値にスコアを振る

(6) スコアが 1 に最も近い値を の最適値とする

$nDCG$  は情報検索システムが生成するランキングの良し悪しを評価する際に用いられる評価指標である。上記の手続きでは、 の成る順位の回数が重複を含み、 $nDCG$  の計算に必要な関連度スコアを、1 位から 20 位のなかで成った順位の回数とみなせるため、この評価指標を用いた。探索の結果上位 5 件の の値と  $nDCG$  のスコアを表 1 に示す。

表 1 より、パラメータ の重みを -0.3 とした。例として、実際に決定した重みで特徴語を抽出した結果を表 2 に示す。

(注12): [http://www.anlp.jp/resource/journal\\_latex/](http://www.anlp.jp/resource/journal_latex/)

表 1 パラメータ の最適値決定のための  $nDCG_{20}$  の結果上位 5 件

$nDCG_{20}$	
0.917	-0.3
0.911	-0.2
0.911	0
0.894	-0.6
0.862	-0.5

表 2 提案特徴語抽出手法で抽出した特徴語上位 20 個

貼るだけ探索:RFID タグの検出履歴を利用した物探し支援システム	マイクロ生体認証:人間の微細生体領域を利用した生体認証	全天球ボールカメラによる撮影動画の視点固定手法の提案
擬似 タグ RFID 距離 物 Tag リーダー 貼付 検出 位置 スキャン 写真 パッシブ 目的 ユーザ 無向 電波 用品 移動 システム	生体 認証 画像 箇 所 利用 肌理 マイクロ 部位 要求 情報 ユーザ 撮影 テンプレート システム マーク 証 微細 可能 マッチング	天球 カメラ 回転 撮影 動画 ボール 姿勢 固定 高緯度 視点 球面 緯度 フレーム 画像 座標 特徴 図法 距円筒 アルゴリズム 領域

#### 4.3 575 候補の生成

Rafal らの研究 [9] では 312 個の俳句を読み、代表的な文法テンプレートを人手で決定し、抽出した特徴語をテンプレートに当てはめることで俳句を生成した。本研究はこれを参考に、575 テンプレートに特徴語を割りあてる方式を採用することとした。

575 テンプレートとは、論文から抽出する特徴語を当てはめることで完成する 575 の骨組みであり、予備実験で準備した学術論文 82 本分の 575 を用いて作成した。具体的には、(1) 575 の一つ一つに形態素解析を行って各単語に品詞を振り、(2) 名詞部分を抽出した特徴語と置換するための記号に変換し、(3) 助詞や接続詞といった機能語部分についてはそのまま機能語を配置し、82 個のテンプレートを用意した。その後、代表的なテンプレートを 20 個人手で選出した。作成した 575 テンプレートの例をいくつか表 3 に示す。

表 3 作成した 575 テンプレートの例

575 テンプレート
<unk><unk> <unk><unk> で <unk> する
<unk><unk> で <unk><unk> <unk> する
<unk> <unk><unk> の <unk><unk>

4.2 節で述べた特徴語抽出により、論文から単語スコアの高い上位 20 個の名詞を抽出する。これを用意した 20 個の 575 テンプレートに対して割り当て、すべての組み合わせ (575 候補) を生成した。575 候補の集合のサイズは、 $20P_6 + 20P_5 \times 12 + 20P_4 \times 6 + 20P_3 = 50937480$  である。なお、575 テンプレート 20 個のうち、名詞を 6 個含むテンプレートが 1 つ、5 個含むものが 12、4 個含むものが 6、そして 3 個含むテンプレートが 1 つ存在する。

#### 4.4 合議制スコアリングによる 575 の決定

生成した 575 候補の中から最終的な出力を決めるために各々に対して評価を行い、最も適切なものを論文要約 575 の出力とした。575 候補の評価は以下に示す合議制スコアリングによって行った。本研究の合議制スコアリングでは次の 4 点について合議をとる。

- タイトルとの対応
- アブストラクトとの対応
- 係り受け関係
- ポジティブ度・ネガティブ度

この 4 点は、予備実験にて学術論文から 575 を人手で作成した際、より良い 575 を決定するために筆者が重視した点を経験則として合議制スコアリングに反映させたものである。具体的には以下の尺度で各 575 を評価する。

- (1) タイトルとの  $\cos$  類似度
- (2) アブストラクトとの  $\cos$  類似度
- (3) 上五・中七・下五の各句の係り受けスコア
- (4) 575 に含まれる単語のポジティブ・ネガティブスコア (ポジネガスコア)

(1) と (2) の  $\cos$  類似度におけるベクトルの計算には単語の出現回数を用いた。(3) の係り受けスコアの導出には、係り受け解析器の CaboCha<sup>(注13)</sup> を用い、575 中の係り受け関係を上五・中七・下五のそれぞれがどれくらい係りあっているかのスコアで取得し、その総和をスコアとした。(4) のポジネガスコアの導出には、2005 年に高村らの研究によって生成された単語感情極性対応表<sup>(注14)</sup> を用いて各単語にスコアを振り、その合計をスコアとした。なお、単語感情極性対応表に存在しない単語にはスコア 1 を振り、単語頻度が多いものは 1 に近く、少ないものは 0 に近づくように出現頻度によって階段がつくようにした。

上記 4 つの尺度を用いた合議制スコア (式 (5)) の高い順に、各 575 をランク付けし、最終的なひとつを決定する。

$$\begin{aligned} \text{合議制スコア} = & * \text{タイトルとの } \cos \text{ 類似度} \\ & + * \text{アブストラクトとの } \cos \text{ 類似度} \\ & + * \text{係り受けスコア} \\ & + * \text{ポジネガスコア} \end{aligned} \quad (5)$$

式 (5) の  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  はパラメータである。

(注13): <http://taku910.github.io/cabocha/>

(注14): [http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_en.html](http://www.lr.pi.titech.ac.jp/~takamura/pndic_en.html)

ここで、各パラメータの重み決定アルゴリズムを述べる。合議制スコアのパラメータは、ひとつの重みが-1.0から1.0までの範囲の値をとるとして、ランダムサーチによる機械学習で最適な組み合わせを見つけた。ランダムサーチのコスト関数には文書要約システムの要約率を測る際に用いられる ROUGE-N [18] と呼ばれる評価指標を用いて計算するシステム 575 (システムが生成する 575) と参照 575 (人手で作成した 575) の N-gram 類似度 (式 (6), 式 (7)) を F 値 (式 (8)) で導出した値を用い、このコストを最大化するパラメータの組み合わせを探索した。  $unigram(C)$  はシステム 575 に含まれる単語集合、  $unigram(R)$  は参照 575 に含まれる単語集合である。  $Count_{match}(e)$  はシステム 575 と参照 575 で一致する単語を数える関数であり、  $Count(e)$  は unigram の出現回数を数える関数である。

ROUGE-N では、単語の N-gram 一致率を用いて 2 つの文章がどれだけ類似しているかを測り、特に人手で作成した要約 (参照要約) とシステムで生成した要約 (システム要約) の類似度を測る。本研究では、ROUGE-N のなかでも ROUGE-1、つまり unigram (1 単語) の一致率を測ることとしたため、式 (6) では、「システム 575 が参照 575 に含まれる単語をどれだけ当てられたか」を表しており、また、式 (7) では「システム 575 の単語が参照 575 にどれだけ含まれているか」を表す。一般的に、Recall と Precision はトレードオフの関係にあり、適切な落とし所を探す必要があるため、双方の調和平均を式 (8) により求めることでバランスのとれた所 (F 値) を求めた。

$$ROUGE_1(Recall) = \frac{\sum_{e \in unigram(C)} Count_{match}(e)}{\sum_{e \in unigram(R)} Count(e)} \quad (6)$$

$$ROUGE_1(Precision) = \frac{\sum_{e \in unigram(C)} Count_{match}(e)}{\sum_{e \in unigram(C)} Count(e)} \quad (7)$$

$$F_{value} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (8)$$

ランダムサーチは以下の手続きで行った。

(1) WISS2016 の論文からランダムに選択した 10 個の論文をデータセットとする

(2) 5 個の論文それぞれを (3)~(6) の処理にかける

(3) 1 個の論文に対してランダムサーチとしてパラメータの重み組み合わせを 1000 個作り (iteration = 1000), (4)~(5) を繰り返す

(4) 重み組み合わせを合議制スコアに適用し、575 候補を合議制スコアの高い順に並び替える

(5) 575 候補上位 20 件それぞれの F 値を求め、総和を当該パラメータのコストとし、前回のコストよりも高ければコストと重み組み合わせを更新する

(6) 最高コスト (F 値の合計が最高) のパラメータの重み組み合わせを保持する

パラメータ探索によって 5 個の論文から 5 つのパラメータの重み組み合わせを得られた。表 4 に示す。

5 つのパラメータの重み組み合わせのうち、どれが最適か決めるためにバリデーションを行った。(1) データセットの残り 5

表 4 ランダムサーチで得られたパラメータの重み組み合わせ 5 つ

組み合わせ 1	0.9	-0.8	0.6	0.0
組み合わせ 2	-0.9	0.8	-0.1	0.1
組み合わせ 3	-0.8	0.1	0.2	-0.6
組み合わせ 4	-0.8	0.8	-0.1	0.6
組み合わせ 5	0.6	0.7	0.3	-0.1

つの論文それぞれに 5 つの重みの組み合わせを全て適用し、(2) 各論文の 575 候補を合議制スコアの高い順に並び替え、(3) 上位 20 件に  $ROUGE_1$  を適用して得られる F 値の和を求め、各論文において F 値の和が最も高くなった回数が最多のパラメータの重み組み合わせを最適解とした。結果を表 5 に示す。

表 5 パラメータの重み組み合わせを他論文に適用したときに F 値の和が最も高くなった回数

	回数
組み合わせ 1	1
組み合わせ 2	0
組み合わせ 3	1
組み合わせ 4	1
組み合わせ 5	2

表 5 より、F 値の和が最も高くなった回数が最多であるのは組み合わせ 5 の [0.6, 0.7, 0.3, -0.1] となった。したがって、合議制スコアリングの各パラメータの重みを  $= 0.6$ ,  $= 0.7$ ,  $= 0.3$ ,  $= -0.1$  とした。

#### 4.5 出力例

学術論文の入力に対する実際の出力の例をいくつか示す。

- 貼るだけ探索：RFID タグの検出履歴を利用した物探し支援システムの提案<sup>(注15)</sup>

検出タグ 貼付の位置に リードする

- マイクロ生体認証：人間の微細生体領域を利用した生体認証<sup>(注16)</sup>
- 認証に 画像と要求 利用する

- 全天球ボールカメラによる撮影動画の視点固定手法の提案<sup>(注17)</sup>
- 動画カメラ ボールと撮影 固定する

## 5. 評価実験

提案手法によって生成する 575 (システム 575) は要件 (「少ない文章量で論文を表現できる」、「印象に残り易い」) を満たしているか確認するために、3 つの実験を行った。

### 5.1 実験 1: 合議制スコアリングの妥当性評価

システム 575 の評価を行う前に合議制スコアリングの妥当性を評価する必要があると考えたため、合議制スコアによって順位付けをする 575 「ソート済みシステム 575」上位 5 件と、ランダム抽出の順位付けをしない 575 「未ソートシステム 575」

(注15): <http://www.wiss.org/WISS2016Proceedings/oral/16.pdf>

(注16): <http://www.wiss.org/WISS2016Proceedings/oral/05.pdf>

(注17): <http://www.wiss.org/WISS2016Proceedings/oral/21.pdf>

5 件を比較する実験を協力者 8 名に行ってもらった。具体的には、DEIM Forum 2017<sup>(注18)</sup> よりランダムに選択した論文 1 本<sup>(注19)</sup> からソート済みシステム 575 と未ソートシステム 575 を生成し、それぞれを混在させた 10 件のシステム 575 を協力者に提示し、「良い」と「悪い」の 2 つのグループに 5 個ずつ分ける試行を行ってもらった。結果は表 6 のとおりとなった。

表 6 ソート済みシステム 575 と未ソートシステム 575 の「良い」評価の数 (最大 5)

	ソート済みシステム 575	未ソートシステム 575
1 人目	1	4
2 人目	3	2
3 人目	2	3
4 人目	3	2
5 人目	4	1
6 人目	3	2
7 人目	5	0
8 人目	1	4
合計	22	18
平均	2.75	2.25

表 6 より、良いグループに選ばれた数はソート済みシステム 575 の方が多い結果となった。また、表 6 は各人がそれぞれのシステム 575 につけた得点 (最大 5 点) と見ることができる。そこで、5 % の有意水準で t 検定を試みたが、有意差は見られなかった。回答に開きが生じた理由として、「良い」と「悪い」の基準は個人の嗜好に依存することが考えられる。しかし、合計得点のみで比較するとソート済みシステム 575 の方が評価が高いため、提案した合議制スコアリングによって 575 候補を適切に順位付けできることとする。

### 5.2 実験 2: 論文を表現できるか確認

システム 575 が論文を表現できているかを確認する実験を 2 つ行った。1 つ目の実験では、実験協力者 7 名に論文のアブストラクトと参照 575 (手製の論文 575) を結びつける問題と、論文のアブストラクトとシステム 575 を結びつける問題合計 10 題に回答してもらった。実験には WISS 2016 よりランダムに選出した 10 本の論文を用いた。各問題 5 題ずつ用意し、それぞれを混在させた状態で提示することで回答者にはシステム 575 か参照 575 か不明なようにした。各問題の選択肢は別の論文から生成したものも含めて 5 つ用意した。本実験では、システム 575 の正解率が参照 575 と同等、もしくはそれ以上であれば論文を表現できていることとした。回答の結果、システム 575 と参照 575 の正解率がともに 100 % となったため、システム 575 で論文の概要を表現できることを確認した。

2 つ目の実験では、上述の実験で扱った論文よりも狭い専門領域における内容の近い論文を提案手法に入力した場合に、システム 575 で各論文を表現できるかを確認する実験を行った。実験協力者 8 名を半分に分け、論文アブストラクトとタイトルを結びつける問題 5 つを前半の 4 名に、後半 4 名には論文アブ

ストラクトとシステム 575 を結びつける問題 5 つに回答してもらった。実験には DEIM Forum 2017 のセッション「レシピ (1, 2)」にある論文からランダムに選択した 5 本を用い、2 つの問題で同じ論文アブストラクトを扱った。実験の結果、双方の正解率は 100 % となったため、同じ専門領域における内容の近い論文についても、システム 575 で細かな違いを汲み取ったうえで論文概要を表現できることを確認した。

まとめると、システム 575 はタイトルと同等の役割を持ち、タイトルよりも少ない文章量で論文を表現できることを確認したため、タイトルよりも一貫性が高くなる結果につながった。

### 5.3 実験 3: 論文が印象に残り易いか確認

システム 575 が印象に残りやすいかを確認するため、論文アブストラクトとそれに対するシステム 575 と参照 575 の 2 択を回答者 7 名に提示し、当該アブストラクトが印象に残りやすい (記憶しやすそう) と感じる方を選択してもらった。このとき、回答者は選択肢のどちらがシステム 575 かは分からない。本実験では実験 2 とは別の論文を 10 本選出し、10 題用意した。回答者は適当に選択した 1 題のみに回答し、回答後、選択した問題のシステム 575 と参照 575 を形容詞対の表に基づいて評価してもらった (Semantic Differential Method, SD 法)。実際の形容詞対の表を表 7 に示す。

表 7 実験 3 に用いた形容詞対の表

	非常に	かなり	やや	どちらでもない	やや	かなり	非常に	
楽しくない								楽しい
落ち着かない								穏やか
窮屈な								ゆったりした
醜い								美しい
劣る								優れる
ごみごみした								整った
よそよそしい								親しみのある
浅い								深い
悪い								良い
複雑								簡素
不快								快適
ぼんやりした								はっきりした
乏しい								豊か
古い								新しい
くだらない								賢い

実験の結果、7 人がシステム 575 よりも参照 575 の方が印象に残りやすそうと回答した。SD 法により回答してもらった 575 の評価を基にシステム 575 と参照 575 の印象評価を行ったところ、7 点スケール (最低 1 点, 最高 7 点) でシステム 575 は平均 4.28 点, 参照 575 は平均 4.97 点となり、双方は 1 % の有意水準で有意差があった ( $p < 0.01$  in t-test)。結果を図 4 に示す。

以上より、システム 575 は手製の参照 575 よりも印象に残りにくい結果となった。特に図 4 より、システム 575 は参照 575 より「劣り、複雑である」印象が強いことがわかる。これは、提案手法で生成する 575 には音韻的な読みやすさはあるものの、単語の不自然なつながりが見受けられることが原因ではないかと考えられる。

## 6. 結 論

本研究では学術論文を読むべきかどうか判断する手間と論文の概要を思い出す手間を軽減することを目的に、575 形式の

(注 18): <http://db-event.jpn.org/deim2017/proceedings.html>

(注 19): <http://db-event.jpn.org/deim2017/papers/243.pdf>

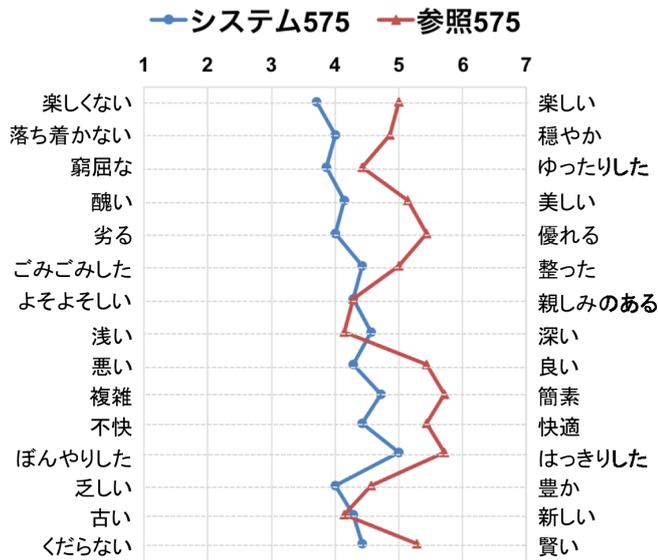


図 4 形容詞対によるシステム 575 と参照 575 の印象の比較 (SD 法)

キャッチフレーズで論文の要約文を生成する手法を提案した。提案手法は「特徴語抽出」と「575 候補生成」、候補絞り込みのための「合議制スコアリング」の 3 つのモジュールで構成される。

本研究のアイデアは、古来より親しまれてきた俳句や川柳といった 575 の音韻的読みやすさを論文の要約文生成に適用すれば、論文を短く表現でき、かつ印象に残りやすくていい点である。そのため、本研究は読者に対して論文を読むべきかどうか判断する手間と概要を思い出す手間を軽減することで貢献し、著者に対しては論文に付与するキャッチフレーズやキーワードの生成を代行、もしくは考案支援での貢献が期待される。

提案手法を評価実験した結果、提案手法により出力される 575 は、少ない文章量で論文タイトルと同様に学术论文の概要を表現することが可能であることを確認した。一方で、印象に残りやすい効果は手製の 575 と比較したところ劣る結果となってしまった。これは、システム 575 には音韻的な読みやすさはあるものの、単語の不自然なつながりが含まれることによる影響と考える。そのため、合議制スコアリングによる評価の際に単語のつながりも加味した基準の考案を今後の課題とする。その他にも、入力論文の全文をそのまま文字列として利用しているため、論文の持つ構造を活用した評価手法を考えることも課題とする。

本研究のシステムが確立すれば、手元にある学术论文を読むべきかどうかスクリーニングする際や、読み返したい論文の概要を想起する際に役立つ。他にも、どんな文章量の論文からでも同じ形式の文を生成するため、学会の査読プロセスにて査読者が論文を一覧する場面や学会の Web ページで参加者が論文を一覧する場面でも役に立つことが期待される。また、17 音程度の文字列であれば空いた隙間に挿入されやすいだろうという考えより、Web 広告などで用いられるフレーズ (謳い文句) の生成に適用できることが期待される。

#### 文 献

[1] 坂野信彦. “七五調の謎をとくー日本語リズム言原論”. 東京, 大修館書店, 2004.9, 273pp.

[2] 渡部涼子 他. “五七調・七五調のリズム知覚に関する予備的研究”. 言語処理学会第 20 回年次大会発表論文集, pp. 574–577, 2014.

[3] 越場千絵 他. “暗記学習のための語呂和歌作成支援システムの開発”. 卒業論文, 熊本大学, 2004.

[4] 中須賀謙吾 他. “談話構造を利用した学术论文の自動要約生成”. 言語処理学会第 21 回年次大会発表論文集, pp. 569–572, 2015.

[5] Amjad Abu-Jbara et al. “Coherent Citation Based Summarization of Scientific Papers”. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pp. 500–509, 2011.

[6] 衣川和亮 他. “学术论文の章構造に基づくニューラル自動要約モデル”. 言語処理学会第 23 回年次大会発表論文集, pp. 150–153, 2017.

[7] 安藤一秋 他. “論文概要からのタイトル自動生成の試み”. 言語処理学会第 10 回年次大会発表論文集, pp. A10C2–04, 2004.

[8] 大部達也 他. “Recurrent Neural Network を用いた抽出型および生成型論文タイトル生成について”. The 31st Annual Conference of the Japanese Society for Artificial Intelligence, pp. 3A1-1, 2017.

[9] Rafal Rzepka et al. “Haiku Generator That Reads Blogs and Illustrates Them with Sounds and Images”. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, pp. 2496–2502, 2015.

[10] Daniel Winterstein et al. “X575: writing rengas with web services”. Proceedings of the INLG 2016 Workshop on Computational Creativity in Natural Language Generation, pp. 27–30, 2016.

[11] Yael Netzer et al. “Gaiku: Generating Haiku with Word Associations Norms”. Proceedings of the NAACL HLT Workshop on Computational Approaches to Linguistic Creativity, pp. 32–39, 2009.

[12] Martin Tsan Wong et al. “Automatic Haiku Generation Using VSM”. Proceedings of the 7th WSEAS International Conference on APPLIED COMPUTER & APPLIED COMPUTATIONAL SCIENCE, pp. 318–323, 2008.

[13] David M. Blei et al. “Latent Dirichlet Allocation”. Journal of Machine Learning Research 3, pp. 993–1022, 2003.

[14] Chong Wang et al. “Online Variational Inference for the Hierarchical Dirichlet Process”. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, pp. 752–760, 2011.

[15] Yee Whye Teh et al. “Hierarchical Dirichlet Processes”. Journal of the American Statistical Association 101, pp. 1566–1581, 2006.

[16] Daniel Gildea et al. “Topic-based language models using EM”. Proceedings of the 6th European Conference on Speech Communication and Technology, pp. 2167–2170, 1999.

[17] Kalervo Jarvelin et al. “Cumulated Gain-Based Evaluation of IR Techniques”. Journal of the ACM Transactions on Information Systems, pp. 422–446, 2002.

[18] Chin-Yew Lin et al. “ROUGE: A Package for Automatic Evaluation of Summaries”. Proceedings of the Workshop on Text Summarization Branches Out, Post Conference Workshop of ACL, pp. 74–81, 2004.