情報処理学会論文誌 Vol.62 No.2 538–547 (Feb. 2021)

重要図表を考慮した End-to-Endの学術論文要約システム

佐藤 照大[1] 中沢 実[1,a)] 松井 くにお[1] 阿部 倫之[1]

受付日 2020年5月12日, 採録日 2020年11月5日

概要：近年，注目度の高い学術テーマでは特に論文投稿数・採択数が激増しており，研究者のサーベイや査読の負担が問題となっている．そこで，文と図表を要約するスライド生成の研究が進められている．しかし，既存研究では文章要約の発展に対し，図表の重要性にはあまり注目されていない．サーベイにおいては要約文だけでなく最適な図表を効率的に把握することは重要である．そこで本研究では，文章要約と図表抽出・選択を組み込んだ End-to-End の要約スライド生成システムを開発した．本システムではユーザが GUI 上で論文を選択することで自動的に要約スライドを生成できる．また，提案システムの有用性を示すため，本研究では各機構の性能評価に加え，ユーザの内容理解度と満足度の評価を行った．その結果，特に内容理解度においては人手で作成したスライドに近い高い正解率を得られた．

キーワード：自動要約，スライド生成，図表抽出

End-to-End Academic Paper Summarization System that Takes into Account Important Figures and Tables

Akihiro Sato[1] Minoru Nakazawa[1,a)] Kunio Matsui[1] Noriyuki Abe[1]

Received: May 12, 2020, Accepted: November 5, 2020

Abstract: In recent years, the number of submissions and acceptance of papers has increased dramatically in high-profile scientific themes, and the survey and peer review burden on researchers has become a problem. Therefore, the generation of slides for summarizing sentences and figures is being studied. However, existing research has not paid much attention to the importance of figures and tables in the development of written summaries. It is important to grasp not only the summary text but also the most appropriate figures and tables efficiently. In this study, we developed an end-to-end summary slide generation system that incorporates text summarization and figure selection. Our system can generate summary slides automatically by selecting an article on the GUI. In order to demonstrate the usefulness of the proposed system, we evaluated the performance of each mechanism as well as the level of understanding and satisfaction of users. As a result, we obtained a high percentage of correct answers, especially in the comprehension of the content, which was close to that of manually created slides.

Keywords: summarization, slide generation, figures extraction

1. はじめに

近年，人々が簡単に得られる情報が増え続け，各々が自力で良質な情報を選別しなければならなくなった．そこで自動的に重要な情報を選別する「自動要約の技術」が注目されている．特にテキストを対象とした自動要約を文章要約と呼び，大きく抽出型と生成型の手法に分類される．抽出型は文書中から直接抜き出す手法であり，柔軟性は低いが原文をそのまま使用できるという利点がある [5], [7], [9], [12]．生成型は柔軟で流暢な文を生成できるが，原文の意図を正しく反映できない可能性がある [4], [17]．Encoder-Decoder モデルの発展により生成型が主流となっていたが，近年

1 金沢工業大学
Kanazawa Institute of Technology, Nonoichi, Ishikawa 921–8501, Japan
a) minoru.nakazawa.jp@ieee.org

© 2021 Information Processing Society of Japan

では事前学習モデル [6] の成功もあり，抽出型と生成型を組み合わせたハイブリッド型の研究がさかんとなっている [11], [22].

自動要約の対象は多様であるが，我々はその中でも学術論文に着目した．学術論文は簡単にアクセスできる機会が増えているだけでなく，人気の高い研究分野では特に論文投稿数・採択数が急増している．たとえば Computer Vision の国際会議である CVPR2019 では前年度に比べて投稿数が約 1.5 倍になっている[*1]．また，採択前に arXiv のようなオープンアクセスリポジトリで公開するケースも多く，こうした論文の注目度も近年では高まっている．このような現状から，特に経験の浅い研究者は研究分野に関する知識が少ないため，効率的なサーベイを行うことが困難となってきている．経験豊富な研究者であってもサーベイに必要な負担は当然増加してしまう．さらには査読者 1 人あたりの負担も非常に重くなるため，査読の質の低下も危惧されている．

論文のサーベイで重要なのは正確かつ素早く内容を把握することであるため，要約スライドにおいて重要な文章と図表を選択することは必須である．しかし，これまでの研究では，要約スライドの自動生成における図表選択についてあまり注目されてこなかった．また，要約スライドで重要なのはスライドからどれだけ正確かつ効率的に内容を理解できたかであるのに対し，理解度の評価が十分にされてこなかった．

本研究では，学術論文のサーベイにかかる負担を軽減することを目的として，論文から自動的に要約スライドを生成するシステムを開発した．ユーザは，**図 1** に示した GUI 上でサーベイしたい論文の PDF を選択し，論文の要約文と重要図表がセットになったスライドを生成できる．本システムでは原文をより正確に反映できる利点から要約文の生成には抽出型を採用した．本システムは要約文と図表を決定する 2 つの行程に分かれており，要約文抽出モジュールでは PDF からテキストを抽出し，各ページに適した要約文を fastText [1] で分類した結果を用いて LexRank [7] により決定する．図表抽出モジュールでは DeepFigures [18] で生成されたページ全体の画像と座標値を用いて図表を抽出し，キャプションに対する TF-IDF とコサイン類似度を用いて各ページに適した図表を選択する．

システム全体の評価のため，11 人の評価者にアンケート調査と理解度調査を行い，スライドの品質を検証した．さらに各モジュールの性能を評価するため，文章要約には自動評価指標の ROUGE，図表選択には採用した図表の一致度を用いて比較・検証を行った．

*1 http://cvpr2019.thecvf.com/files/CVPR%202019%20-%20Welcome%20Slides%20Final.pdf

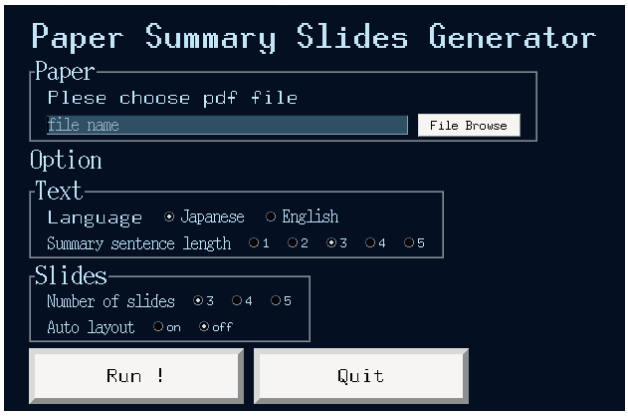


図 1 ユーザが使用するシステムの GUI
Fig. 1 GUI used by the user.

2. 関連研究

2.1 文章要約

文章要約は大きく抽出型と生成型に分類される．抽出型では文や単語の類似度から抽出する文を決定する手法が提案されている．Erkan ら [7] は PageRank [2] をもとにノードを文，エッジを類似度としたグラフベースの手法である LexRank を提案した．また，Word2Vec などの Embedding を用いた手法 [9] も提案されている．ニューラルネットワークを活用した研究もされており，Denil ら [5] は CNN でラベル付けされた文書から関連する文を抽出する手法を提案した．Nallapati ら [12] は Encoder-Decoder を拡張したモデルで CNN Daily Mail Dataset[*2] を学習させ，各文の抽出確率を出力する SummaRuNNer を提案した．抽出型の利点は，原文から直接文章を抜き出すことで原文の意図を正確に要約文に反映できることである．一方で原文にない表現を自由に使えないため，文間を自然につなぐことができないという欠点を持つ．

生成型ではニューラルネットワークを活用した研究が特にさかんである．Rush ら [16] は Attention 付きの Encoder-Decoder モデルで生成型要約を行った．しかし，系列変換モデルには未知語（OOV）に弱く情報の繰返しが起こるという欠点があった．たとえば学習データで使用されていない単語が入力文に登場した場合，Unknown を表す一律の ID が振られてしまう．これに対して Yao ら [21] は OOV に対処するため Decoder の入力文から語彙をコピーする CopyNet を提案した．See ら [17] は生成とコピーの両方を行う生成型要約手法を提案した．See らの手法では，Decoder の Attention の分布を利用する Coverage Mechanism を加えることで重複を制御している．生成型の利点は文間を考慮した自然な文章を生成できることにある．しかし生成された文章の内容の正確性は保証されておらず，意図に反して虚偽の内容を生成してしまう恐れがあ

*2 https://github.com/deepmind/rc-data

© 2021 Information Processing Society of Japan

る [3].

近年では Transformer [19] を用いた事前学習モデルの登場により，転移学習を用いて複数のタスクで高性能を誇るモデルが提案されている [6], [14]．また，文章要約タスクに特化した事前学習モデルも提案されており，高精度な要約が生成可能であると報告されている [11], [22].

2.2 学術論文を対象とした自動要約

学術論文を対象とした自動要約では，Abstract を参照要約とする研究のほかに，要約スライドを生成する研究がさかんである．柴田ら [25] は生のテキストに談話構造解析を行い，主題部と非主題部をスライドに箇条書きで記述した．

学術論文にはテキストだけでなく内容を補う図表が存在する．サーベイの目的は論文の内容を理解することであるため，生成するスライドに図表を採用することは重要である．安村ら [24] は TeX 形式の論文・ページ数・キーワードを入力として XHTML 形式のプレゼンテーション資料を生成するシステムを開発した．Yamamoto ら [20] は YOLOv2 [15] でタイトル・著者・Abstract・図表を取得し，単一ページの要約を生成する手法を提案した．Yamamoto らの研究では手動で注釈をつけたデータセットを作成し学習させることで，論文の PDF から各バウンディングボックスを認識している．谷口ら [26] は fastText [1] の分類結果から，学術論文を効率的にまとめる方法として知られる落合フォーマットの形式で単一ページを生成する手法を提案した．谷口らの手法では fastText の分類結果が高いものを重要文と定義しており，文全体の重要度は加味していない．また，入力は自動化されていないほか，スライドに含める図表は Poppler [*3]で抽出した最初の画像を採用している．

2.3 図表抽出とレイアウト生成

学術論文の図表は内容を把握するための重要な情報が含まれている．しかし，ラベル付きデータセットが存在せず，構造化されたテキストを使用するか [24]，自前で注釈を付与したデータセットを作成する必要があった [20]．Siegel ら [18] は arXiv と PubMed の補助データを利用して大規模なデータセットを作成した．具体的には学術論文の LaTeX または XML に図表を認識しやすいよう注釈を付与し，550 万個以上のラベルを含んだデータセットを作成した．さらに ResNet101 [8] を使用し，単一画像に変換した論文から図表を抽出する DeepFigures を提案した（**図 2**）．DeepFigures を使用することで arXiv，PubMed に投稿された論文に対して F1 score で 80%以上の精度を達成できたと報告されている．

また，作成者が効率的にデザインを決定できるよう，自動的にレイアウトを生成する手法も提案されている．Qiang ら [13] はグラフィカルモデルを利用してポスタの要素となる各パネルの属性や配置を推論するフレームワークを提案した．Zheng ら [23] は画像を VGG16，キーワードを Word2Vec，カテゴリを One-hot ベクトルでそれぞれ埋め込む特徴量抽出ネットワークと，Generative Adversarial Network（GAN）を組み合わせてレイアウトを生成する手法を提案した．

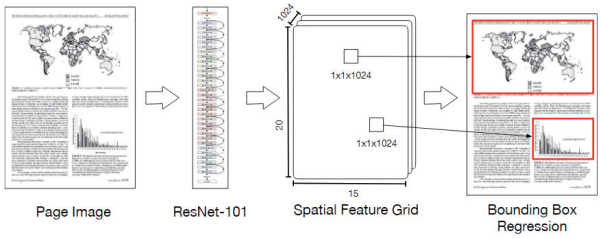


図 2 DeepFigures のモデル概要 [18]
Fig. 2 DeepFigures model.

3. 提案システム

本研究では，学術論文のサーベイにかかる負担を軽減することを目的とした要約スライド生成システムを開発する．ユーザは GUI 上でサーベイしたい学術論文を選択することで，自動的に要約スライドを生成できる．本システムの概要を**図 3** に示す．提案システムは要約文抽出モジュールと，図表抽出モジュールの 2 つの機構で構成されている．本研究で生成するスライドは，入力する論文の長さにかかわらず，サーベイに必要最低限な情報であると考えた「タイトル」「概要」「新規性・結果」「コメント・考察」の 4 枚とした．本研究では，テキスト分類において日本語のコーパスを作成・学習させるため，日本語で書かれた学術論文を要約対象とする．

3.1 要約文抽出モジュール

本システムでは学術論文の PDF を入力とする．まず pdfminer [*4]を使用して PDF からテキストを抽出する．pdfminer の抽出結果には不要な文字列が多く含まれるため，注釈や参考文献を含む不要な文を前処理で除去し，テキスト分類のための fastText へ入力する．fastText は subwords という構成要素単位で分割することで Word2Vec よりも高い精度で Embedding を実現した手法である [1].

本システムにおけるテキスト分類の目的は，タイトルを除く 3 つのページに最適な文を振り分けることである．コーパスを作成するため，cvpaper.challenge が公開している CVPR2018 の日本語まとめスライド 942 本のコンパイラである Pugs [*5]を利用した．タグなどの不要な文字を削

*3 https://poppler.freedesktop.org/

*4 https://pypi.org/project/pdfminer/
*5 https://github.com/cvpaperchallenge/CVPR2018_Survey

© 2021 Information Processing Society of Japan

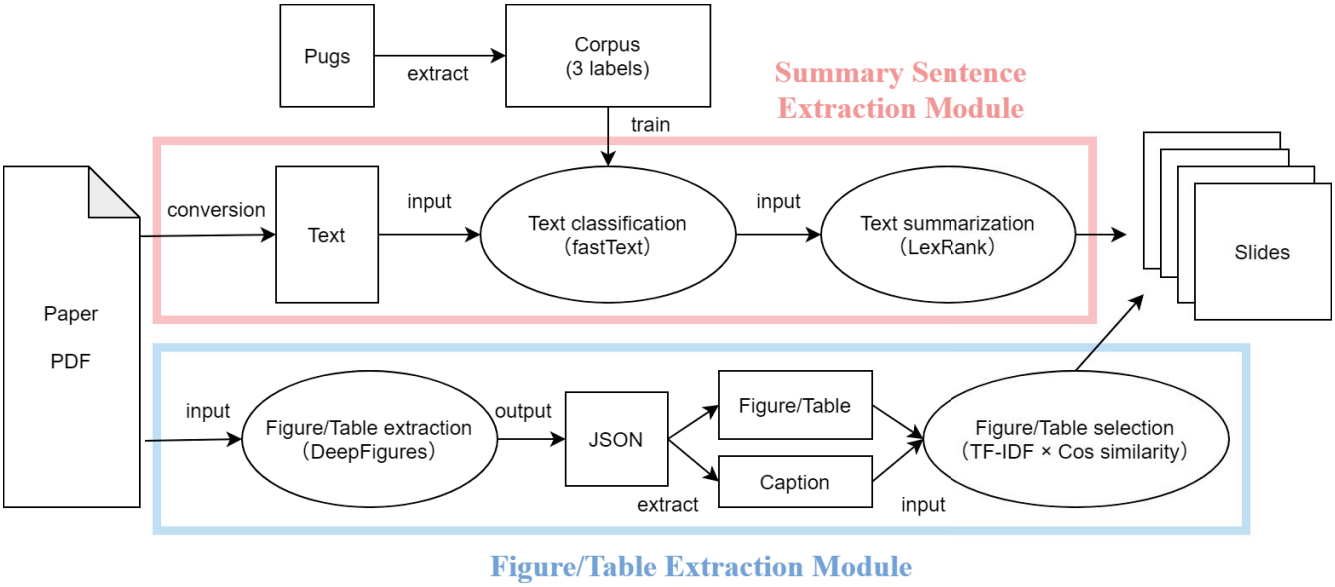


図 3　提案システムの概要
Fig. 3　System overview.

除した後，MeCab [*6]により分かち書きしたテキストから 3 種類のラベル付きコーパスを作成した．コーパスはランダムに 8:2 で分割し，学習・検証に使用した．システム動作時には fastText によって学術論文に含まれる各文を「概要」「新規性・結果」「コメント・考察」の 3 つのカテゴリに分類した後，文章要約によって重要文を決定する．本研究では要約文の内容の正確性を重要視し，抽出型の LexRank [7] を採用した．実装には gensim [*7]を利用している．

3.2　図表抽出モジュール

本研究では図表抽出に DeepFigures [18] を用いた．提供されている検証用の実装[*8]をもとにライブラリの依存関係を修正したものを使用し，arXiv と PubMed で学習された重みで学術論文の PDF から各ページの画像と，図表の座標値を推定した JSON を生成する．続いて JSON から図表の座標値とキャプションを取得し，画像に変換された各ページから図表を抽出する．

次に要約スライドに採用すべき重要図表を決定する．DeepFigures でキャプションの座標値を取得できているため，重要文との類似度が高いキャプションを重要度が高い図表と定義し，重要図表を選択した．ただし DeepFigures で認識できるのは英語のみであるため，googletrans [*9]によるキャプションの翻訳を行っている．DeepFigures で取得した座標値をもとに併記されている日本語キャプションを取得することも可能だが，座標値の推定精度と OCR などによる文字列の抽出精度が課題となるため，本研究では単純なライブラリによる翻訳を採用している．また，類似度計算には googletrans の翻訳精度による影響とノイズとなりうる品詞の削除を目的として，MeCab で取得した名詞のみを使用している．

類似度計算の手法としてコサイン類似度および 4 つの分散表現獲得手法（TF-IDF，Doc2Vec [10]，Word2Vec の平均，fastText の平均）をそれぞれ検証した結果，本研究では最も単純かつ性能が期待できる TF-IDF とコサイン類似度を採用した．検証した手法のうち，Doc2Vec は実行時に類似度が大きく変動する傾向があった．これは対象としているキャプションの文長が短すぎることや，データ不足が影響していると考えられる．またドメイン特有の学習済みデータを用意することは困難であるため Doc2Vec を除外した．Word2Vec や fastText の平均を使用することで類似度を計算できるが，学術論文に登場する OOV の対応が難しい．たとえば近年の自然言語処理の研究では BERT などの固有名詞が使用されるが，学習データに存在しない単語の重要度は正しく推定できない．OOV の対処としては無視や追加学習が考えられる．しかし学術論文のテーマは多岐にわたり，Doc2Vec と同様の問題が生じる．学術論文においては OOV となる情報こそが重要なキーワードとなりうると考え，これらの手法も除外した．

本システムでは古典的であるがドメインに影響されない TF-IDF とコサイン類似度を用いて，要約文とキャプションの 2 文間の類似度計算を行い，重要図表を選択した．また，図表はスライドの各ページに抽出された要約文ごとに類似度を計算し，最も高い類似度を優先的に掲載するよう処理を行っているため，同一の図表が選択されることはない．また，要約文抽出においても同様に，文章が重複した

*6　http://taku910.github.io/mecab/
*7　https://radimrehurek.com/gensim/
*8　https://github.com/allenai/deepfigures-open
*9　https://pypi.org/project/googletrans/

© 2021 Information Processing Society of Japan

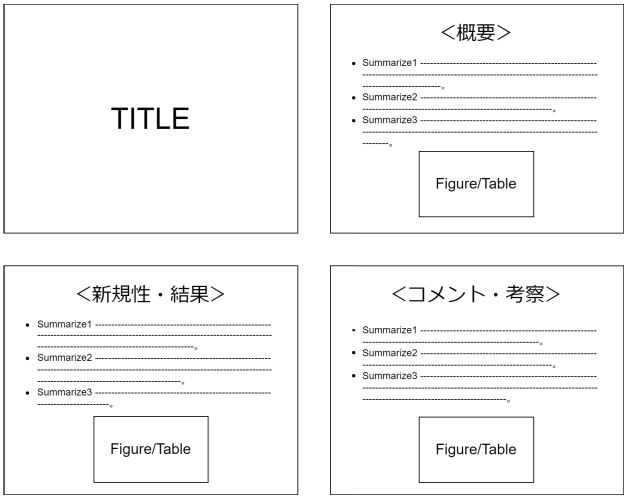


図 4　生成されるスライドの固定レイアウト
Fig. 4　Layout of the generate slides.

際には次に優先度の高い文章を選択するようにしている．

3.3　レイアウト構成

本システムで生成されるスライドのレイアウトは固定とした．本来であればユーザが自由にデザインを変更できることが望ましいが，本実験では要約文と重要図表の抽出に着目し，レイアウト生成は今後の課題とした．

本研究では PDF のファイル名をタイトルとして 1 ページ目に掲載し，残りの 3 ページにはそれぞれ要約文が 3 文と，図表が 1 枚掲載できる固定のレイアウトを作成した（**図 4**）．スライドの生成には python-pptx [*10] を利用し，pptx 形式のスライドを出力する．各スライドのレイアウトは固定の座標値で決定しているため，文章量や図表の縦横比によってはレイアウトが崩れてしまう．そこで本システムでは各パネルの位置と図表サイズを固定し，文章量によってフォントサイズを変動させた．

4.　評価実験

評価実験のために情報処理学会論文誌のオープンアクセス可能な論文 5 本を対象に，本システムを用いて要約スライドを生成した．また本システムの比較対象として，2 人の大学院生に同様に要約スライドを作成してもらった（人手スライド）．この 2 人の作成者には 5 本の論文のうち，できる限り専門分野に近い論文からスライドを作成できるよう，1 人の作成者には 2 本を，もう 1 人の作成者には残りの 3 本を担当してもらった．大学院生を人手スライドの作成者とした理由は，論文にある程度読み慣れていることを理由にスライドの品質を保証できるためである．また，作成者を 1 人にする方が作成基準が明確になるが，疲労により品質の低下を招く恐れがある．逆に 5 人にすると目標となる人手スライドの基準が曖昧となるため，本実験では 2 人

[*10] https://python-pptx.readthedocs.io/en/latest/

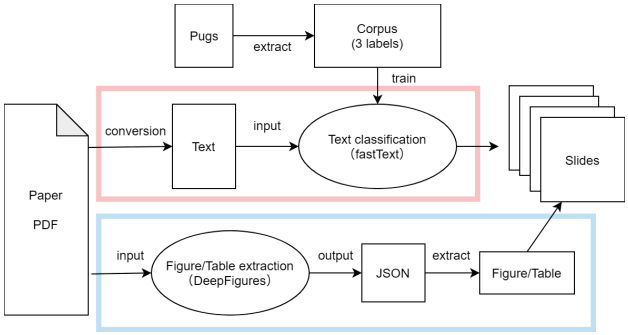


図 5　比較用のシステム．fastText のコーパスと DeepFigures による図表抽出を固定し，文章要約の除去と図表選択の変更を行った
Fig. 5　A system for comparison. We fixed the corpus of fastText and the figure extraction by DeepFigures, removed the text summary and changed the figure selection.

が妥当な人数であると判断した．人手スライドの作成に要した時間の平均は 1 スライドあたりおおよそ 45 分であった．本実験ではデザインを評価対象外としているため，2 人の作成者には図 4 をもとにスライドを作成してもらった．また評価者が見た目で判別できないようにするため，要約文は原文から 3 文ずつ抜き出し，図表は 0〜3 枚使用してもらった．

提案システムの比較として，谷口ら [26] と同様に，fastText の分類結果を要約文として採用するシステムを構築した．しかし谷口らの手法では先頭に登場する図表 1 枚を採用しており，提案システムと異なり図表の重要度が考慮されておらず，選択図表の枚数も少ない．そのため，提案システムが評価時に有利となる可能性がある．そこで本研究においての提案内容の価値を明確化するために，我々の提案システムから要約文生成機構を取り除き，図表選択の方法として登場順で 3 つの図表をスライドに掲載するよう変更したものを比較システムとして構築した（**図 5**）．さらに，fastText の分類結果のみを使用すると，前処理で除去しきれなかった不要な文字列や，pdfminer で変換できなかった文字列による影響が大きい．そこで比較システムのみ 20 字以下の文と，英数字記号のみで構成される文を除外した．

本研究では学術論文 5 本を対象に，人手・提案システム・比較システムの 3 つの手法でスライドを作成した．スライドはタイトルを除き 3 ページで構成されており，各ページには要約が 3 文，図表が 1 枚使用されている．ただし人手スライドに限り，図表の採用有無は作成者に委ねている．すなわち各手法で生成されたスライドのページ数は各 15 ページであり，図表の数は最大で各 15 枚である．

本研究ではテキスト分類・文章要約・図表選択の各機構を自動評価指標を用いて性能を検証し，さらに，人手による内容理解度クイズとアンケートによる生成スライドの主観評価を行った．

© 2021 Information Processing Society of Japan

表 1　本タスクにおける fastText の性能評価
Table 1　Evaluation of fastText performance in this task.

	50 slides	942 slides
Accuracy	0.556	0.603
F1 score (macro)	0.550	0.565
F1 score (weighted)	0.557	0.602

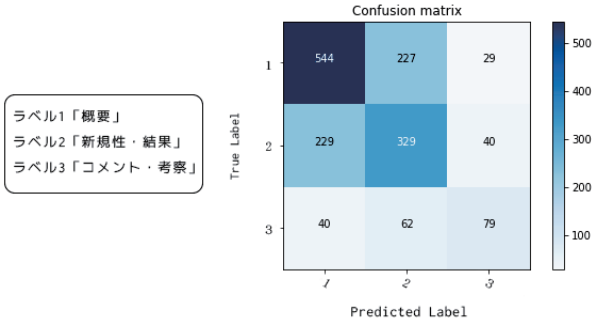


図 6　分類結果より得られた混同行列
Fig. 6　Confusion matrix obtained from the classification results.

なお，本実験は Intel Core i9-9900KF CPU および GeForce RTX 2080 を搭載した Ubuntu 18.04 上で実行している．10 ページの論文を対象に本システムを 5 回実行したときの平均動作時間は 34.71 秒であった．ただし GUI 上で論文を選択する時間は除いている．また，実行速度は DeepFigures に大きく依存し，DeepFigures を除いた平均実行時間は 10.95 秒であった．

4.1　fastText による分類結果

942 本の論文まとめスライドで作成したコーパスをランダムに 8:2 に分割し，fastText の性能を検証した．訓練データのうちラベル 1「概要」が 3,280 個，ラベル 2「新規性・結果」が 2,390 個，ラベル 3「コメント・考察」が 642 個であった．ここで 50 本のみをランダムに抽出し学習・推論したものを比較として用いた．**表 1** に fastText の分類結果，**図 6** に混同行列を示す．Accuracy と F1 score の計算には scikit-learn を用いた．macro はマクロ平均，weighted は加重平均である．

分類結果より，本システムではおよそ 6 割の精度で分類できることが分かった．また，データ数が多いほど，さらに性能が向上する可能性も示唆されている．しかし増加したデータ数に対して性能向上が乏しい．これは使用した学習データが複数人によって作成されたものであり，各ラベルの認識と品質にブレが生じていたことが要因であると考えられる．すなわち，クラウドソーシングのような大人数での学習データの作成は，ルールを厳密に固定しない限り，容易に分類性能の上昇が期待できない可能性がある．

図 6 より，「概要」は比較的高い精度で分類できているが，「コメント・考察」の分類精度は低いことが分かる．これは人によって考察を端的に書いたり，別のリンクで補足したりするなど，品質の差が大きかったためではないかと考えられる．また，最も枚数が少ないのもラベル 3「コメント・考察」であった．そのため学習自体がうまくいっておらず，前述のように高品質なデータを増やせば全体の分類精度も大きく上昇する可能性がある．さらに，今回よりもラベルの種類を増やすことで分類のルールが厳密になり，性能の向上が期待できる．

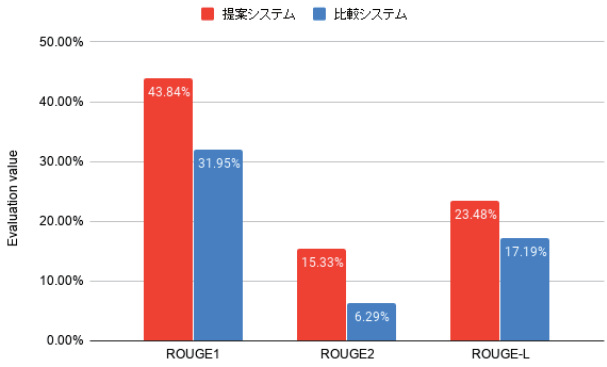


図 7　ROUGE による要約文の評価結果
Fig. 7　Results of ROUGE evaluation.

本研究では統一された執筆基準でのラベル付きコーパスを作成するのは困難と考え，既存のデータを活用することでラベル付きコーパスのコストを削減した．一方で，統一された執筆基準がないために品質が低下し，分類結果に大きく影響してしまったと考えられる．本研究では谷口ら [26] の手法をもとに論文の章構造を使用せず，読解に必要な 3 種類のラベルを採用した．しかし論文中における「コメント・考察」に該当する記述には執筆基準がなく，そもそもコーパスの品質にかかわらず分類が困難である可能性は否定できない．そのため今後の課題として，今回採用した 3 種類のラベルが分類方法として最適かを検証していく必要がある．

4.2　要約文と図表選択の自動評価

生成した要約文を評価するために自動要約の評価指標として一般的な ROUGE スコアを使用した．本研究では ROUGE-N と ROUGE-L を使用する [10]．**図 7** に人手スライドに対する評価結果を示す．

提案システムの評価値は，比較システムよりも ROUGE-1 で 11.89%，ROUGE-2 で 9.04%，ROUGE-L で 6.29%上昇し，提案システムの方がすべてのスコアで上回った．すなわち End-to-End のシステムにおいて，fastText の分類結果を採用するよりも文章要約を追加する方が高品質な要約文を抽出できる．

続いて人手スライドで使用された図表と，各システムが選択した図表の一致度を検証した．異なる図表であっても内容が類似している場合があり，理解を助ける可能性があ

© 2021 Information Processing Society of Japan

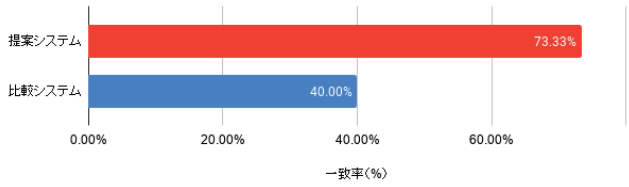


図 8　選択図表の一致率による評価

Fig. 8　Evaluation by the agreement rate of the selected figure and table.

る．しかし，類似の判定は主観によるものになってしまい，適切な評価が困難となる．そこで本研究では人手スライドに掲載された図表と同一の図表を選択された場合のみを一致とした．ただし，ページ単位で図表を選択すると論文全体として重要でありながらも選択されない可能性が生じるため，掲載ページにかかわらず人手スライドで使用されている場合には一致と判定している．また，人手スライドのうち 1 ページのみ作成者が図表を採用しないという選択肢をとった．本来であれば文と図表が一致していなければ理解の妨げになることもある．しかし，担当した作成者は複数の図表が理解の助けになる可能性があったが，選択が困難だったため採用しなかったと提言していることから，本実験では該当スライドの図表 1 枚のみどんな図表であっても加点要素として計算している．一致率の比較を**図 8** に示す．図表抽出においても提案システムの生成スライドの方が，比較システムの生成スライドよりも評価が上回った．つまり，登場順で図表を選択するよりも，キャプションと要約文の類似度を計算することで，重要な図表を選択できると考えられる．

4.3　クイズによる手動評価

4.1 節，4.2 節では各構造の性能について評価したが，最も重要なのはどれだけユーザの理解に役立つ要約スライドを生成できるかということである．そこで人手・提案システム・比較システムの 3 種類の方法で生成されたスライドをもとに，内容の理解度調査を行った．クイズの内容は google フォームを用いて一般公開している[*11]．評価者として大学生・大学院生の 9 人に協力してもらった．評価者を A・B・C の 3 グループに無作為に分割し，1 人あたり 5 本のスライドを読んで各クイズに回答してもらった．対象となる計 15 本のスライドは**表 2** に示すように無作為に各グループに振り分けた．また，評価者にはスライドを作成したのが誰か，システムが生成している可能性も含めて伝えていない．

クイズは各スライドの生成前に作成した．クイズの内容は論文の核心にできるだけ近く，問題として答えやすいものを基準に記述式・選択式の両方の形式で作成した．クイズの数は各論文に対して 5 個ずつ出題し，1 人あたり計 25 個のクイズに回答してもらった．ただし正当な評価を行うため，評価者にはスライド以外を参照して回答することを禁止した．記述式の採点は論文中の記述と一致しているかを基準に行い，論文内でまったく記されていない回答は不正解とした．選択式はすべて 4 択で出題している．

*11 クイズの内容：https://forms.gle/sLTmtJRKx9HU7toP7

表 2　無作為に各グループに振り分けたスライド一覧

Table 2　List of slides randomly assigned to each group.

	Slide 1	Slide 2	Slide 3	Slide 4	Slide 5
人手	A	B	C	C	B
提案	C	C	A	B	A
比較	B	A	B	A	C

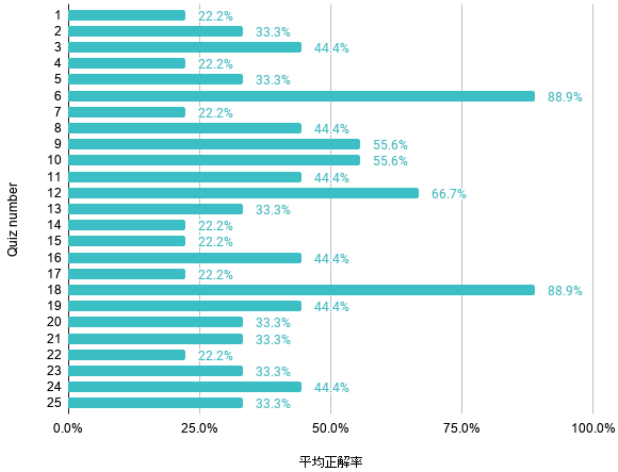


図 9　各クイズの平均正解率

Fig. 9　Average percentage of correct answers for each quiz.

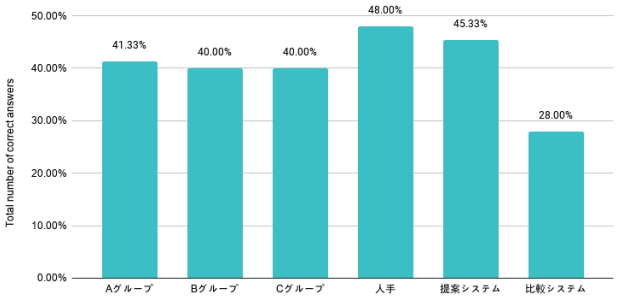


図 10　グループと手法ごとの正答率の比較

Fig. 10　Comparison of the percentage of correct answers by group and method.

図 9 に各クイズの平均正解率を示す．すべてのクイズの平均正解率は 40.4%となった．選択式の方が正答数が低く，記述式が 41.1%，選択式が 37.8%であった．

図 10 に各グループごとの正答数と，各手法ごとの正答数の比較を示す．グループごとの正答数では，グループによって正答数に差がほとんどなく，評価者によって回答の品質にばらつきが少ないことが分かった．手法ごとの正答数では正答数に開きがあり，人手スライドが最も高く 48%，続いて提案システムが 45.33%，比較システムが最も低く 28.0%であった．人手スライドは論文を読み慣れた大学院

© 2021 Information Processing Society of Japan

表 3　実施したアンケート内容
Table 3　Details of the questionnaire conducted.

Q1	「概要」を読んで内容を理解できましたか？
Q2	「新規性・結果」を読んで内容を理解できましたか？
Q3	「コメント・考察」を読んで内容を理解できましたか？
Q4	適切な画像が使用されていると思いましたか？
Q5	このスライドは論文を理解するために役に立ちますか？

生が作成したものであり，要点がまとまった高品質な要約スライドであると考える．また，人手スライドは各システムと同等の条件で作成されているため，各システムと人手スライドの正答率を比較することで，どれだけ人手に近い高品質なスライドを生成できたかの検証が可能となる．結果より，提案システムの正解率は人手スライドにわずかに及ばなかったが，その差は 2.67%と僅差であり，提案システムが内容理解度の点で高品質なスライドを生成できたことが明らかとなった．すなわち，提案システムよって論文の要点がまとまった要約スライドの自動生成が可能となり，研究者の負担の軽減を実現した．

一方で，人手スライドの内容理解度が 48.0%にとどまっていることから，より理解度が高くなるようなレイアウトの検証が必要である．ただしサーベイで必要なのは素早く効率的に内容を把握することであるから，クイズの正解率が高いほど良いスライドであるとは限らない．そのため，サーベイの速度と正答率の均衡を検証していく必要がある．

4.4　アンケートによる主観評価

もう 1 つの人手による評価として，アンケートを実施した．アンケートの内容は google フォームを用いて一般公開している[*12]．評価者には，結果に影響がないよう人手スライドの作成者と 4.3 節で述べた評価者とは異なる大学生・大学院生 3 人に協力してもらった．アンケートの内容は**表 3** に示した 5 つの質問を各スライドに対してそれぞれ行い，5 段階評価（1 が最も低く，5 が最も高い）で回答してもらった．クイズによる評価と同様に，スライドの作成者は伝えていない．各評価者には 15 本すべてのスライドを読んでもらい，回答を集計した．また，図表が使用されていないページに限り，図表が使用されていないことが適切か判断してもらった．

図 11 に手法別に集計したアンケート結果を示す．結果より，5 つのアンケート内容すべてにおいて比較システムよりも提案システムが生成したスライドの方が評価が高くなった．4.3 節で述べたとおり，人手スライドに近い評価が得られたかどうかで生成スライドの品質が明らかになるが，提案システムの生成スライドとの評価結果には差が開いてしまった．つまり，内容理解度では僅差となったにもかかわらず，ユーザの感情的には満足を与えることができ

*12 アンケートの内容：https://forms.gle/GHRHCSkyVjYsLoBE9

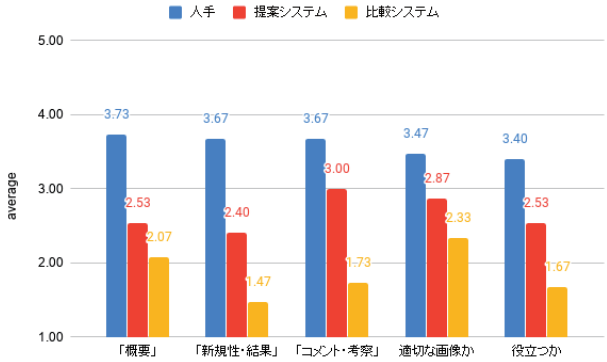


図 11　手法別に集計したアンケート結果
Fig. 11　Questionnaire results by method.

なかった．これは自動生成によって生じた誤字・脱字や，要約文どうしの文脈構造を考慮できなかったことが大きな原因ではないかと考えられる．

各項目に注目すると，最も提案システムの評価が高かったのは「コメント・考察」のページである．これは「～と考える」などの理解しやすいキーワードを含んだ文をうまく抽出できたのではないかと考える．最も提案システムの評価が低かったのは「新規性・結果」のページである．これは論文ごとに主張が異なるため他のページよりも文脈依存が大きかったと考えられる．

「適切な画像か」という質問はおおむね評価が高い．この評価結果は 4.2 節の一致度と同様の傾向であり，キャプションによる図表の重要度決定は自動要約スライドの満足度に影響するといえる．すなわち要約スライドにおいて，要約文だけでなく図表選択に着目することは重要である．

4.1 節と同様に，本研究では既存のデータを活用しているため，主観評価においても統一された基準を設けていないことの弊害が少なからず生じている．執筆基準を厳密に定義し，ラベル付きコーパスを作成することで評価結果の向上が期待できるが，反対にコーパスの作成コストが課題となる．そこで今後の課題として，論文の章構造を本手法における分類ラベルに採用し，高品質かつ作成が容易なコーパスの検討をしていく必要がある．

5.　結論と今後の課題

本研究では学術論文のサーベイにかかる負担を軽減することを目的として，要約文抽出モジュールと図表抽出モジュールからなる要約スライド生成システムを開発した．ユーザには最低限の操作でシステムが扱えるよう GUI アプリケーションを提供し，End-to-End のスライド生成を実現した．本システムに 10 ページの学術論文を入力したときの 5 回の平均動作時間は，GUI 上で論文を選択する時間を除いて 34.71 秒であった．人手スライドは平均 45 分で作成していることから，速度の面では明らに人手に勝っている．また，生成スライドは直接論文を読む代わりとし

© 2021 Information Processing Society of Japan

て使用できるだけでなく，ユーザの目的に応じて文章量や図表枚数を調節できるよう拡張が可能であるため，従来のアブストラクトに代わる論文全体を俯瞰するための用途としても使用できる．

評価には人手・提案システム・比較システムの 3 種類の手法を用いた．fastText による分類結果では，提案システムが 60.3%の精度で分類できることを提示し，より高い精度を達成するためには高品質な学習データを作成する必要性があることを示した．ROUGE と画像の一致率による評価では，いずれも提案システムによる評価が比較に対して優れており，テキスト分類後に文章要約機構を追加することで高品質な要約文を抽出できることを検証した．また，図表抽出においては DeepFigures で抽出した座標値をもとに図表を抽出し，キャプションに対して TF-IDF とコサイン類似度で計算を行い，比較システムよりも正確に重要図表を選択できることを検証した．さらにスライド全体の品質を評価するため，人手による 2 種類の評価を実施した．クイズによる内容理解度評価では，9 人の評価者を 3 つのグループに振り分けて実施し，提案システムが人手に近い高品質なスライドを生成できることを示した．アンケートによる主観評価では，提案システムの生成スライドは比較システムよりも評価が高い一方で，人手スライドとは差が開いており，機械的にうまく処理できなかった文字列が悪影響を及ぼした可能性が明らかとなった．

本研究ではユーザのために GUI を提供したが，システムの利便性やレイアウトの妥当性の検証はしていない．また，既存研究においても要約生成システムのユーザビリティには注目されておらず，ユーザ目線での評価が不足しているのが現状である．そこで今後の課題として，システムの利便性の評価や効率的なレイアウトの検討が必要である．文章要約では，近年成功している事前学習モデル [6], [11], [22] を採用し，より柔軟な要約文を生成することでより高い理解度・満足度が得られると考えられる．また，入力段階において誤認識した文字や，前処理で除去できなかった文字列を検証し，より高精度で入力テキストを整形する機構の追加が必要である．図表選択ではキャプションを利用する特性上，文長に依存してしまう．そこで原文と図表参照位置を考慮した図表選択手法を検討していく必要がある．

参考文献

[1] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T.: Enriching word vectors with subword information, *Trans. Association for Computational Linguistics*, Vol.5, pp.135–146 (2017).

[2] Brin, S. and Page, L.: The anatomy of a large-scale hypertextual web search engine (1998).

[3] Cao, Z., Wei, F., Li, W. and Li, S.: Faithful to the original: Fact aware neural abstractive summarization, *32nd AAAI Conference on Artificial Intelligence* (2018).

[4] Chen, Y.-C. and Bansal, M.: Fast abstractive summarization with reinforce-selected sentence rewriting, arXiv preprint arXiv:1805.11080 (2018).

[5] Denil, M., Demiraj, A. and De Freitas, N.: Extraction of salient sentences from labelled documents, arXiv preprint arXiv:1412.6815 (2014).

[6] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[7] Erkan, G. and Radev, D.R.: Lexrank: Graph-based lexical centrality as salience in text summarization, *Journal of Artificial Intelligence Research*, Vol.22, pp.457–479 (2004).

[8] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[9] Kobayashi, H., Noguchi, M. and Yatsuka, T.: Summarization Based on Embedding Distributions, *Proc. 2015 Conference on Empirical Methods in Natural Language Processing*, pp.1984–1989, Association for Computational Linguistics (online), DOI: 10.18653/v1/D15-1232 (2015).

[10] Lin, C.-Y.: ROUGE: A Package For Automatic Evaluation of Summaries, *ACL 2004* (2004).

[11] Liu, Y. and Lapata, M.: Text summarization with pretrained encoders, arXiv preprint arXiv:1908.08345 (2019).

[12] Nallapati, R., Zhai, F. and Zhou, B.: Summarunner: A recurrent neural network based sequence model for extractive summarization of documents, *31st AAAI Conference on Artificial Intelligence* (2017).

[13] Qiang, Y., Fu, Y., Guo, Y., Zhou, Z.-H. and Sigal, L.: Learning to Generate Posters of Scientific Papers, *AAAI* (2016).

[14] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer, arXiv preprint arXiv:1910.10683 (2019).

[15] Redmon, J. and Farhadi, A.: YOLO9000: Better, faster, stronger, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.7263–7271 (2017).

[16] Rush, A.M., Chopra, S. and Weston, J.: A neural attention model for abstractive sentence summarization, arXiv preprint arXiv:1509.00685 (2015).

[17] See, A., Liu, P.J. and Manning, C.D.: Get to the point: Summarization with pointer-generator networks, arXiv preprint arXiv:1704.04368 (2017).

[18] Siegel, N., Lourie, N., Power, R. and Ammar, W.: Extracting scientific figures with distantly supervised neural networks, *Proc. 18th ACM/IEEE on Joint Conference on Digital Libraries*, pp.223–232 (2018).

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I.: Attention is all you need, *Advances in Neural Information Processing Systems*, pp.5998–6008 (2017).

[20] Yamamoto, S., Fukuhara, Y., Suzuki, R., Morishima, S. and Kataoka, H.: Automatic paper summary generation from visual and textual information, *International Conference on Machine Vision* (2019).

[21] Yao, T., Pan, Y., Li, Y. and Mei, T.: Incorporating copying mechanism in image captioning for learning novel objects, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.6580–6588 (2017).

[22] Zhang, J., Zhao, Y., Saleh, M. and Liu, P.J.: PEGASUS:

© 2021 Information Processing Society of Japan

Pre-training with Extracted Gap-sentences for Abstractive Summarization, arXiv preprint arXiv:1912.08777 (2019).
[23] Zheng, X., Qiao, X., Cao, Y. and Lau, R.W.: Content-aware generative modeling of graphic design layouts, *ACM Trans. Graphics* (*TOG*), Vol.38, No.4, pp.1–15 (2019).
[24] 安村禎明，武市雅司，新田克己：論文からのプレゼンテーション資料の作成支援，人工知能学会論文誌，Vol.18, No.4, pp.212–220 (2003).
[25] 柴田知秀，黒橋禎夫：談話構造解析に基づくスライドの自動生成，自然言語処理，Vol.13, No.3, pp.91–111 (2006).
[26] 谷口航平，濱川　礼：PTOS：機械学習を用いた論文要旨スライド自動生成システムの提案と開発，情報処理学会インタラクション 2019 論文集，pp.635–640 (2019).

佐藤 照大　（学生会員）

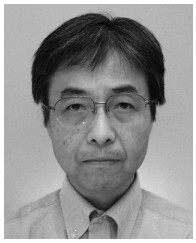
2019 年金沢工業大学工学部情報工学科卒業．現在，金沢工業大学大学院工学研究科情報工学専攻博士前期課程在学．

中沢 実　（正会員）

1993 年金沢工業大学大学院工学研究科情報工学専攻博士前期課程修了．同年（株）富士通研究所入社．1996 年金沢工業大学助手．2011 年同大学教授となり，現在に至る．博士（工学）．自律分散システム，ロボティクス，信号処理等に関する研究に従事．2016 年情報処理学会山下記念研究賞受賞．電子情報通信学会，IEEE 各会員．

松井 くにお　（正会員）

1980 年静岡大学工学部情報工学科卒業．同年（株）富士通研究所入社．2003 年東京工業大学大学院情報理工学研究科後期課程修了．2017 年金沢工業大学教授，現在に至る．博士（工学）．自然言語処理，文書情報処理，情報検索の研究開発に従事．言語処理学会評議員，厚生省電子カルテ研究班班員を歴任．特定非営利法人言語資源協会副会長，日本規格協会 ISO/TC37 国内対策委員．2006 年より本会理事．



阿部 倫之　（正会員）

1989 年金沢工業大学大学院工学研究科情報工学専攻博士後期課程満期退学．同年（株）日立製作所入社．1991 年金沢工業大学情報工学科助手，1998 年同大学助教授，2006 年起業．2012 年同大学准教授となり，現在に至る．工学博士．自律分散システム，SNS 解析等に関する研究に従事．電子情報通信学会会員．

© 2021 Information Processing Society of Japan