

# Word2Vecを用いた挿絵自動挿入手法の提案

村山 貴志<sup>1,a)</sup> 入江 英嗣<sup>1,b)</sup> 坂井 修一<sup>1,c)</sup>

概要：文書のわかりやすさは読み手の理解の早さと正確さにつながりコミュニケーションを円滑化するが、そのためには書き手が様々な点に配慮する必要があることから、すべての書き手がわかりやすい文書を作成することは期待できない。そこで我々は既存の文書をわかりやすくするために挿絵を自動的に挿入する手法を提案する。入力として受け取った任意の長さの文章に対し、その文章の挿絵として最も適切な画像を挿絵候補画像データベースから検索して提示する形となる。検索には画像を説明するテキストを用い、Word2Vecによる単語分散表現で表された距離を複数単語に拡張する手法を利用する。入力文章と最も近い距離にある説明テキストの画像を求める挿絵画像とした。提案手法による挿絵挿入を人手による挿絵挿入と比較する主観評価実験を行い、提案手法には性能面で課題が残ることがわかった。

## 1. はじめに

コミュニケーションの媒体として、文書は大きな役割を占める。例えば企業や政府機関といった組織は、議事録や業務報告で活動を記録し、マニュアルによって組織の新たな構成員に共有知を伝達し、報告書という形で自らの活動を外へと示す。これらはすべて文書という媒体を用いた組織内の構成員同士、あるいは組織内と組織外とのコミュニケーションである。

しかし、わかりやすい文書を書くのは簡単ではない。文書のわかりやすさは読み手の理解の早さと正確さにつながりコミュニケーションを円滑化するという利点がある。そこで多くの書き手はわかりやすい文書を書こうと試みるが、そのためには様々な点に配慮する必要がある。例えば小林 [1] は「知的障害者、高齢で読解機能が低下した人、日本語が堪能でない外国人等、情報弱者の人々」を念頭においてわかりやすい文章のガイドラインを提案しているが、その項目は30にも渡る。さらには福沢 [2] は情報の受け取り手や情報伝達の文脈が変化すれば求められる「わかりやすさ」は変化することを指摘し、質的に異なる様々な「わかりやすさ」の研究を紹介している。このような難しさがある中で、すべての書き手がわかりやすい文書を作成することは期待できない。

そこで我々は、挿絵を自動的に挿入することで既存の文書を読み手にとってわかりやすい文書に作り変える手法を

提案する。タイトル付き挿絵候補画像データベースを事前に用意し、Word2Vec [3] の単語分散表現を用いて文章と各画像タイトルの距離を比較する。文章との距離が最も近い画像をその文章にとって最も適切な挿絵画像であると判断して挿入する。さらに、挿入された画像が適切であるかどうかを主観評価によって確認する。

## 2. 関連研究

### 2.1 挿絵の効果

島田・北島 [4] は、挿絵の文章理解促進効果を説明するために、読解プロセスを「読解初期」「読解最中」に分ける「2段階モデル」を提唱し、それぞれの段階に対応する挿絵の「動機付け効果」と「精緻化効果」を実験により明らかにした。読解初期である2秒間ほどに、読者は少ない情報から読むかどうかの判断を行うと考えられる。このときに挿絵が「読解への方向づけを促進する効果」を動機づけ効果と呼ぶ。読解最中に読者は内容理解のために文書から必要な情報を取得する。このときに「挿絵が心象的表象の形成を促す」効果を精緻化効果と呼ぶ。挿絵の動機付け効果と精緻化効果それぞれについて被験者実験が行われ、実際に存在することが確認された。

石橋 [5] は文章読解において挿絵が存在する場合、その後の記憶保持において優れていることを明らかにした。この研究の本来の目的は読み手が挿絵を自ら描いた場合に長期記憶がより多く保持されることを明らかにすることであったが、実験の対照群として文章のみを読む群と挿絵付きの文章を読む群が設定され、その二者の間で挿絵付きの文章を読む群の方が統計的に有意に長期記憶をより多く保

<sup>1</sup> 東京大学大学院情報理工学系研究科

<sup>a)</sup> murayama@mtl.t.u-tokyo.ac.jp

<sup>b)</sup> irie@mtl.t.u-tokyo.ac.jp

<sup>c)</sup> sakai@mtl.t.u-tokyo.ac.jp

持することが明らかにされた。

## 2.2 挿絵自動生成

五十嵐・上岡 [6] [7] は小説に挿絵を挿入する手法を提案している。この手法は難解なシーンを挿絵によって可視化することで読みやすくすることを目的としているものの、TF-IDF による重要な主人公登場シーンの特定や、シーン背景記述の解析など、小説であることを前提とした処理が多く含まれるため、本稿で対象とするコミュニケーションのための文書をわかりやすくするという目的には利用できない。

## 3. 提案：挿絵自動挿入手法

既存の文書をわかりやすくするために挿絵を自動的に挿入する手法を提案する。提案手法の全体像としては、入力として受け取った任意の長さの文章に対し、その文章の挿絵として最も適切な画像を挿絵候補画像データベースから検索して提示する形となる。

検索対象となる挿絵候補画像データベースは画像とそのタイトルまたは説明文のペアから成り、事前に構築する必要がある。以後この画像のタイトルまたは説明文を「画像テキスト」と呼ぶ。クリエイティブ・コモンズライセンス<sup>\*1</sup>の普及や様々なロイヤリティフリー素材がウェブ上で公開されているという昨今の情勢を鑑みると、多くの挿絵候補画像を有する挿絵候補画像データベースを構築することは難しくない。そして他者に利用されることを念頭に公開されている素材画像には、その画像が何者であるかを記述する画像テキストが設定されていることが一般的である。この素材画像と画像テキストをペアとして、挿絵候補画像データベースを構築する。

挿絵を挿入したい文章のまとまりが入力となり、挿絵候補画像データベースからこの入力に最適な画像を提示したい。そのためには入力文章と画像テキストの類似度が最も高い画像を求めればよい。この類似度を Word2Vec による単語分散表現を拡張することで定義する。単語分散表現では、ベクトルとして表現された単語間のコサイン類似度を計算することで、単語間類似度を 0 から 1 の間で求めることができる。Song & Roth は単語間類似度が与えられているときの二文章間の類似度計算方法として Maximum Alignment という手法を提案している [8]。これは 2 文のうち片方の文に登場するそれぞれの単語について相手の文の単語中で最も類似度が高い単語とのコサイン類似度を計算し、すべての単語についてそれぞれ足しあわせた後に単語数で割ることによって文間類似度を計算するという手法である。計算される値は 2 文について対称でないので、両方の文の側からそれぞれ計算した結果を平均した値

が最終的な文間類似度値となる。しかし、この Maximum Alignment は本研究のように文の長さが著しく異なるような文同士、かつ長い方の文の長さに幅があるような場合において問題が発生する。長い文側から短い文側への類似度計算に際してスコアの低い単語が大量に発生し、ノイズが乗りやすくなってしまう。これを回避するために、本研究においては短い文側から長い文側への類似度計算の値のみをもって類似度と定義する。以後この類似度計算方法を「片側 Maximum Alignment」と呼称する。片側 Maximum Alignment によって入力文章に対する画像テキストの類似度が研鑽されるため、これをデータベースの全件に対して計算し、その中で類似度が最も高い画像テキストとペアとなっている画像を挿絵候補として出力する。これにより文章から最適な挿絵を検索することができる。

## 4. 実装

提案手法を Ruby on Rails<sup>\*2</sup> とその OR マッパーである ActiveRecord を介した SQLite<sup>\*3</sup> データベース操作、および Python の gensim ライブラリ上の Word2Vec を用いて実装した。Ruby のバージョンは 2.5.1、Python のバージョンは 3.4.8 である。

挿絵候補画像データベースはフリー素材提供サイト「いらすとや」<sup>\*4</sup>の素材を用いた。提供されているサイトマップ<sup>\*5</sup>に従いスクレイピングしてデータベースを構築した。スクレイピングは 2019 年 11 月 13 日より開始し、合計 32527 件の画像と画像テキストのペアが作成された。「いらすとや」側に設定された画像テキストはすべて「のイラスト」という文言を含むが、ノイズとならないようデータベース作成時点ですべて取り除いた。

Word2Vec の日本語訓練済みベクトルは鈴木ら [9] の「日本語 Wikipedia エンティティベクトル」2017 年 2 月 1 日版<sup>\*6</sup>を利用した。これは日本語 Wikipedia<sup>\*7</sup>の記事から学習された 200 次元の単語分散表現である。

Song & Roth の Maximum Alignment では、無関係な単語同士のノイズを除去するために単語間類似度 0.5 以下は 0 とみなすという実装上の工夫が存在する。本研究の片側 Maximum Alignment においてもこれを採用し、コサイン類似度 0.5 以下の単語ペアをすべて 0 とみなして計算した。

<sup>\*2</sup> <https://rubyonrails.org/>

<sup>\*3</sup> <https://www.sqlite.org/index.html>

<sup>\*4</sup> <https://www.irasutoya.com/>

<sup>\*5</sup> <https://www.irasutoya.com/sitemap.xml>

<sup>\*6</sup> [http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki\\_vector/](http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/)

<sup>\*7</sup> <https://ja.wikipedia.org/>

<sup>\*1</sup> <https://creativecommons.org/>

## 5. 評価

### 5.1 主観評価

提案手法がわかりやすい文書の作成という目的を達成する手法であるかを確かめるため、主観評価実験を行った。挿絵の存在しない文書に本手法により挿絵を挿入した場合と人間が挿絵を挿入した場合のわかりやすさを比較した。

### 5.2 実験条件

被験者は情報系の大学生または大学院生の7人であった。挿絵の挿入対象文章として、内閣サイバーセキュリティセンターの「サイバーセキュリティ 2019」\*8から1部1章2「変わりゆくサイバー空間とそれに伴う脅威の深刻化」、および外務省の「外交青書 2019」\*9から第1章2「日本外交の展開」を選択した。これらの見出し構造を維持した Markdown 形式のファイルを用意し、挿絵の挿入対象文書とした。

### 5.3 実験手順

被験者7人を4人と3人の2群に分ける。以下前者をA群、後者をB群とする。実験は事前準備と2段階にわたって行われる。

事前準備として、挿絵の挿入対象文書である「サイバーセキュリティ 2019」および「外交青書 2019」からの Markdown 形式抜粋文書それぞれに対し、提案手法によって挿絵を挿入する。挿絵を挿入する部分は、前回挿絵を挿入した直後から1000文字が経過した場所の段落終了直後とする。挿絵と挿絵の間の1000文字以上の文をすべて入力とし、挿絵画像を得る。Markdown の画像挿入記法によって挿絵画像を Markdown 文書内に埋め込み、そのファイルをテキストエディタ Visual Studio Code\*10 の拡張である Markdown PDF\*11 によって PDF 形式に変換する。以上が事前準備である。

実験の第1段階において、被験者は提示された Markdown 形式の文書に挿絵候補画像データベースの画像を用いて挿絵を挿入する。本研究の実装では「いらすとや」が提供する素材画像のスクレイピングにより挿絵候補画像データベースが構築されているため、被験者はスクレイピングが実行された日時以前に投稿された「いらすとや」の素材画像を用いて挿絵挿入を行った。被験者に提示される Markdown 形式の文書は、A群では「サイバーセキュリティ 2019」、B群では「外交青書 2019」からそれぞれ抜粋したものである。被験者によって挿絵が挿入された Markdown 文書は、

事前準備と同様に Visual Studio Code の Markdown PDF 拡張を用いて PDF 形式に変換される。

次に第2段階において、被験者は挿絵が挿入された文書のわかりやすさを評価する。A群の被験者には事前準備と第1段階によってそれぞれ挿絵の挿入された「外交青書 2019」のペア、B群の被験者には同様の「サイバーセキュリティ 2019」のペアが提示される。どちらの文書がどのような方法で挿絵が挿入されたものかは被験者には通知されない。文書は PDF ファイル名により与えたが、ファイル名にランダムな文字列を含めることにより全被験者で先に読まれる文書が固定化することを避けた。被験者は提示された文書のうちどちらがわかりやすいと感じたか、およびその理由を回答する。理由の記述を求めるのはランダムな回答を防ぐためである。

### 5.4 結果

事前準備における提案手法の適用によって、「サイバーセキュリティ 2019」に16件、「外交青書 2019」に11件の挿絵が挿入された。例えば「サイバーセキュリティ 2019」の最初の1000文字区切りは、「新たなサイバーセキュリティ戦略の位置づけ」という表題から始まり、以下の内容である。

新戦略は、基本法に基づき、サイバーセキュリティに関する施策を総合的かつ効果的に推進するために策定した旧サイバーセキュリティ戦略（2015年9月閣議決定。以下「2015年戦略」という。）を初めて改定したものであり、基本法に基づく2回目の「サイバーセキュリティに関する基本的な計画」である。

（中略）

Society 5.0 は、サイバー空間とフィジカル（実）空間を高度に融合させることにより、経済的発展と社会的課題の解決を両立する「人間中心の社会」である。これまでの情報社会では、インターネット上の膨大な情報から必要な情報を人の手で収集・分析するのに限界があったが、Society 5.0 は、人工知能・IoT の活用で、情報の共有・分析が進展し、分野を越えて、人々に様々な恩恵をもたらす社会になるとされる<sup>2</sup>。言い換えれば、次に目指す社会は、サイバー空間と実空間の間でデータが循環して、相互に作用し、これを前提とした様々なサービスが提供され、人々の生活に浸透し、恩恵をもたらす社会である。

この区切りには「走る人たちのイラスト（社会人男性1）」\*12が最も近い画像として挿入された。

\*8 <https://www.nisc.go.jp/active/kihon/pdf/cs2019.pdf>

\*9 <https://www.mofa.go.jp/mofaj/gaiko/bluebook/2019/html/index.html>

\*10 <https://code.visualstudio.com/>

\*11 <https://marketplace.visualstudio.com/items?itemName=yzane.markdown-pdf>

\*12 [https://1.bp.blogspot.com/-XeLDe3y1SIY/XWS5pVa7TjI/AAAAAAAAABUVI/VpLH\\_IIPkA8PaiGVCrr7sY0nJmuIp-2qQCLcBGAs/s1600/kakedasu\\_](https://1.bp.blogspot.com/-XeLDe3y1SIY/XWS5pVa7TjI/AAAAAAAAABUVI/VpLH_IIPkA8PaiGVCrr7sY0nJmuIp-2qQCLcBGAs/s1600/kakedasu_)

主観評価の結果としては、A 群、B 群ともに全員が被験者による挿絵の挿入をより読みやすいと判断した。表 1 に所属していた群と判断の理由をまとめる。

表 1 実験結果

群	理由
A	関連性の高いイラストに直感的な理解を助けられたため。
A	内容に関連したイラストになっていたため、段落ごとの大体の意味をイラストを頼りに理解することができていたから。
A	ex2_eSuDjF <sup>*12</sup> は ex2_hCfXhw <sup>*13</sup> よりも挿絵が多く、とつきやすい印象をうけたから。また、ex2_eSuDjF では挿絵の内容が文章の内容と関連しており、挿絵をとっかかりとして文章を読むことができたから。加えて、ex2_eSuDjF は ex2_hCfXhw と比較して話題と話題の区切りに挿絵が配置されている傾向があり、文章のどこをひとかたまりとしてとらえればよいか分かりやすかったから。
A	- ex2_FYNSLe <sup>*13</sup> において関係のない画像が含まれておりそれが理解の邪魔になったため - 2つのファイルとも文章は同じであり ex2_FYNSLe の方を先に読んだため - ex2_fdXPDz <sup>*14</sup> を読む際には、それに含まれる画像を一度読んだことのある文章の概要をなぞるために使うことができたため。画像を使って内容を思い出すことができた。
B	挿絵と文章の意味内容とは関連があるという暗黙の前提の下読み始めたため、ex2_gnKheH <sup>*16</sup> の絵と文章が一見無関係なことに戸惑った。また、何か関係は無いかと考え始めてしまい、文章の理解とは関係ない思索に時間を費やした。関連は無さそうだと結論したあとにも次々目に入る絵のイメージが頭に残ってしまい、続く文章の理解が妨げられるように感じられた。ex2_uGSNiN <sup>*17</sup> については挿絵と文章の意味内容に関連がありそのような混乱が無かった。また、挿絵を見て下部に書かれた文章の主題をある程度想像してから読み進めるため理解がスムーズだったように思われる。
B	イラストが文章と関係あるから。
B	イラストと文書に関連性があったから

## 6. 結論

既存の文書をわかりやすくするために挿絵を自動的に挿入する手法を提案した。事前に画像と画像テキストのペアからなる挿絵候補画像データベースを構築し、任意長の文章を入力として受け取った際に片側 Maximum Alingment で最も類似度の高い画像テキストの画像を検索し、最適な挿絵画像として推薦する。主観評価により人手での挿絵挿

入と比較したところ、わかりやすさの向上には課題が残ることがわかった。

今後は類似度計算手法や挿絵挿入に最適な文章の区切りについて、より理解の補助となるような挿絵候補画像を選択できるようなアルゴリズムを開発する必要がある。さらに評価指標についても、現状では挿絵の動機付け効果と精緻化効果を区別できていない。被験者の選択理由には「挿絵が多く、とつきやすい印象をうけた」といった動機付け効果を示唆する内容のほか、「関係のない画像が含まれておりそれが理解の邪魔になった」といった精緻化効果の阻害に関する内容が含まれている。この両効果を区別するような主観評価の方法を開発していきたい。

## 参考文献

- [1] 小林美津江：わかりやすい文章作成のための表記法の素材の研究 ソーシャルインクルージョンの社会をめざして，日本保健医療行動科学会雑誌= Journal of the Japan Academy for Health Behavioral Science, Vol. 33, No. 1, pp. 57-67 (2018).
- [2] 福沢将樹：文章・文体(理論・現代), 日本語の研究, Vol. 14, No. 3, pp. 77-82(オンライン), DOI: 10.20666/nihongonokenkyu.14.3.77 (2018).
- [3] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings* (Bengio, Y. and LeCun, Y., eds.), (online), available from (<http://arxiv.org/abs/1301.3781>) (2013).
- [4] 島田英昭, 北島宗雄：挿絵がマニュアルの理解を促進する認知プロセス, *教育心理学研究*, Vol. 56, No. 4, pp. 474-486 (オンライン), DOI: 10.5926/jjep1953.56.4.474 (2008).
- [5] 石橋 薫：挿絵イラストの自己生成が文章内容の記憶再生に及ぼす効果に関する研究, *人間科学研究*, Vol. 25, No. 1, pp. 129-129 (2012).
- [6] 五十嵐晃, 上岡英史：シナリオの内容を反映した挿絵の生成手法, *技術報告 5*, 芝浦工業大学大学院理工学研究科, 芝浦工業大学大学院理工学研究科 (2014).
- [7] 五十嵐晃, 上岡英史：シナリオの内容を反映した場所情報の可視化手法, *SIG-ALST*, Vol. B4, No. 03, pp. 1-6 (2015).
- [8] Song, Y. and Roth, D.: Unsupervised sparse vector densification for short text similarity, *NAACL HLT 2015 - 2015 Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics (ACL), pp. 1275-1280 (2015).
- [9] 鈴木正敏, 松田耕史, 関根 聡, 岡崎直観, 乾健太郎：Wikipedia 記事に対する拡張固有表現ラベルの多重付与, *言語処理学会第 22 回年次大会*, pp. 797-800 (2016).

suit1.png

\*12 被験者により挿絵を挿入された「外交青書 2019」  
 \*13 提案手法により挿絵を挿入された「外交青書 2019」  
 \*14 提案手法により挿絵を挿入された「外交青書 2019」  
 \*15 被験者により挿絵を挿入された「外交青書 2019」  
 \*16 提案手法により挿絵を挿入された「サイバーセキュリティ 2019」  
 \*17 被験者により挿絵を挿入された「サイバーセキュリティ 2019」