

BERT で得られる分散表現に対してコサイン類似度を用いた 小説の挿絵推薦

Recommended Illustrations for Novels Using Cosine Similarity to the BERT based Distributed Representation

下窪 聖人¹ 黄 潤和²

Masato Shimokubo¹, Runhe Huang²

¹法政大学大学院 情報科学研究科

¹Graduate School of Computer and Information Sciences, Hosei University

²法政大学 情報科学部

² Faculty of Computer and Information Sciences, Hosei University

Abstract: In recent years, e-book provides a lot of opportunities that people read books during a commute or break time. E-book reader has some useful functions for reading the books, for instance, Kindle can look up the meaning of words for a reader, and change the font size if necessary. This system enables a function of recommending illustrations during one is reading, which is achieved by calculating cosine similarity to the BERT based distributed representation. There are 4 steps: (1) this system divides text into some scenes by TopicTiling; (2) it extracts the important sentences by BERT for each scene; (3) it extracts named entity using MeCab and CaboCha; and (4) the system recommends appropriate illustrations by calculating similarity between the sentences after this analysis and image title. As expected, this system can understand the contents of the reading easily with recommended illustrations when one is difficult to understand e-book in only text content. The recommended illustrations are original from the illustrations of the web site “irasutoya”. As a result of the experiments, the accuracy of the illustration’s recommendation was about 72.0 percent. This result was obtained by using CNN to convert the image styles of both the original and the recommended illustrations, and then evaluating them by SSIM.

1 はじめに

1.1 研究背景と目的

近年、電子書籍を通勤時間や休憩時間に活用する機会が増えている。このような場面において、読書をしながらか用語を調べたり、知り合い同士で議論し合ったりすることがある。それらをサポートするために、電子書籍リーダーには読書支援ツール機能がある。例えば、小説の単語の意味を調べる機能、テキストを解析して小説の話の流れを可視化する機能などが挙げられる。これらの機能は自然言語処理分野で多く提案されている。しかし、小説を対象とした研究における共通の課題点として、テキストの文脈を高精度で正確に読み取ることが難しい点が挙げられている。この課題点を解決するために多く研究がなされているが、目立った実績はあげられていない。

本研究では、各場面の挿絵を推薦することによって、ユーザの読書をサポートするシステムを提案する。本システムは場面分割をする TopicTiling[1]、重要文を抽出するための BERT[2]を用いる手法[3]、文章解析のための構文解析器 CaboCha[4]と形態素解析器 MeCab[5]を使用し、Web サイト「いらすとや」[6]のイラストを出力する。本システムの出力結果は、シンプルな挿絵を推薦することを目標とする。本研究におけるシンプルな挿絵とは、複数の物体や人物などが描かれている挿絵の場合でも、その挿絵の中で最も重要である物体や人物に注目したイラストと定義する。本システムの有効性を示すことができれば、挿絵の少ない小説に対して挿絵を与えることでユーザの読書をサポートできるようになる。また、書籍の購入時に内容が分からないものがある。そのときに本システムを利用することで、出力から書籍の内容を読み取る手助けになることが期待できる。

1.2 関連研究

読書支援ツールについて様々な研究がある。例えば知的ナビ機能のついた電子書籍リーダを西岡ら[7]は提案している。この研究では、場面分割の代表的手法である TextTiling[8]と登場人物を対象にした固有表現抽出を使うことで場面を解析する。現在の電子書籍リーダの読書支援ツールは、単一のキーワードをクエリとした検索は可能であるが、複数のキーワードをクエリとした検索はできないことが課題点として挙げられている。そのため、場面解析を行う提案システムを利用することで課題点を解決できることを示している。またこの課題点を解決することによって、場面ごとに複数の登場人物の出現を検索することが可能になることを示している。

また、村井ら[9]は、ユーザに向けた電子小説の選別を支援するサービス「立ち読み」インタフェースを提案している。このインタフェースは、未読小説のキーワード検索は難しいという課題点を、その小説のレビューとテキスト両方を解析して可視化することで解決している。具体的には、ユーザの興味を引く単語に対して、出現頻度で重要度を求める手法 TF-IDF[10]を用いて興味喚起度を予測し出力するインタフェースである。このとき、単語の興味喚起度を求める手法は、出現頻度が高くなればなるほど、興味喚起度も比例して高くなるようにしている。また、物語の進行における興味を引く度合いの推移を可視化するために、興味喚起度の変化を層として表現している。

推薦システムが普及していく中で Web 小説に關係する推薦システムの例として、小坂ら[11]の研究を挙げる。小坂ら[11]の研究はユーザの作品検索をサポートする推薦システムである。この推薦システムは作品の類似度を示す S スコアと作品の質を示す Q スコアを用いて推薦を行なっている。このときに Doc2Vec[12]で計算する。既存の推薦システムでは解決できていない読者の少ない小説が推薦されにくいという課題点を解決している。Web 小説を用いた実験によって、推薦システムの有効性を示している。

2 提案システム

本システムは、登場人物とその人物の行動を挿絵として示すことでユーザの読書をサポートする。本システムの構成を図1に示す。本システムの処理の流れは、テキストを図1の手順で解析し、その解析結果によって Web サイト「いらすとや」[6]の画像を出力する。小説「浦島太郎」のテキストを入力として与えると図2のようになる。

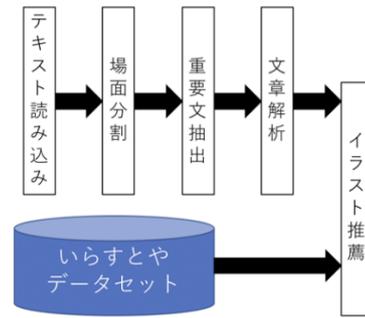


図1 挿絵推薦システムの構成。

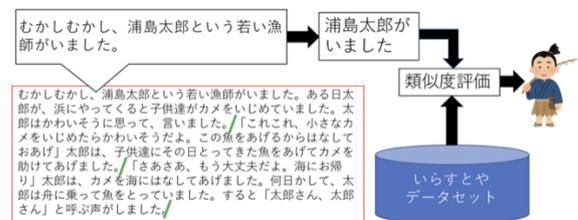


図2 小説「浦島太郎」を入力として与えた結果。

2.1 場面分割

小説や新聞は、読者の理解をサポートするために挿絵を挿入している。挿絵のないテキストのみの小説の挿絵を推薦するためには、テキストを解析して場面を分割し、各場面に対して挿絵を挿入すると読者の理解をサポートできると考える。テキストを場面ごとに分割するときは、作成したベクトルからそれぞれ区切り候補を求める必要がある。一般的に場面を分割するアルゴリズムとして、TextTiling[8]と TopicTiling[1]が使われている。TextTiling[8]は、単語間の類似度と閾値を比較することで分割を行う。この手法の課題点は、1文が短い文章が入力として与えられると、周辺の単語数が少ないことにより適切に場面を分割できないことが知られている。そこで、単語の出現情報を使用しない手法として、トピックモデルを使用する TopicTiling[1]が提案されている。TopicTiling[1]は、トピックモデルの Latent Dirichlet Allocation (LDA) [13]を使用してトピックベクトルを作成し、場面境界の判定に使用する depth score を求めている。LDA[13]は、トピックごとに単語の出現確率の生成や、文章ごとにトピック確率の生成に使用されるモデルである。TopicTiling[1]の depth score は以下の式で求め、 p は位置、 c_p はある位置の coherence score、 $hl(p)$ は p より左側にある範囲で coherence score の最大値を返す関数、 $hr(p)$ は p より右側にある範囲で coherence score の最大値を返す関数である。本システムでは、1文が短い場合にも対

応できる TopicTiling[1]を使用する。また、coherence score はコサイン類似度で求める。

$$\text{depth score} = \frac{1}{2(hl(p) - c_p + hr(p) - c_p)} \quad (1)$$

2.2 重要文抽出

挿絵は各場面の重要な文をイラストとして表していると仮定すると、各場面の文章には内容と関係が薄い文が存在する可能性がある。重要文抽出によってこの文を取り除くことで、読者にとって必要な挿絵を推薦できると考える。重要文抽出は Yang Liu ら [4]が提案した手法を改良して使用する。Yang Liu ら [4]は、初めて文章要約タスクに BERT[3]を適用し、ニュース記事のデータセットで高精度の要約を実現した。Yang Liu ら [4]が提案した手法は英語のみに対応しているため、日本語の小説を処理できるように前処理を加える必要がある。本研究では、形態素解析器 MeCab[5]を使って分かち書きしたテキストを入力として与える。また、日本語の学習済みモデルとして日本語の Wikipedia を学習した京都大学黒橋・村脇研究室のモデル[14]を使用する。

2.3 文章解析

文章解析では、登場人物は主語として必ず出現すると仮定する。このように仮定した理由は、小説は登場人物の行動や発言を中心に話を進める傾向があると考えたからである。本システムの文章解析は、各場面の重要文に対して構文解析と形態素解析を行う。これらの解析により、主語を key、目的語を含む述語を value とする辞書を出力する。例として「うらしまたろうはカメに乗った」を入力すると、辞書 {うらしまたろう : カメに乗った} が出力される。処理の流れとして、まず構文解析で主語と述語を抽出する。処理中にある文の主語が見つからない場合は、直前の主語をその文の主語とした。次に、形態素解析を行い、抽出された主語が適切であるかを判定する。主語が適切でない判定された場合、その key を削除する。そして、形態素解析の処理で残った辞書を出力する。

2.4 イラスト推薦

文章解析後の辞書と Web サイト「いらすとや」[6]のデータセットのラベルとの類似度を比較して、最も類似度が高いラベルを探索する。類似度を求める時には、BERT[2]で得られる分散表現とコサイン類

似度を使用する。「いらすとや」[6]のデータセットから、このラベルを持つイラストを出力する。この出力が本システムの出力結果となる。Web サイト「いらすとや」[6]のデータセットは公開されていないため、スクレイピングによってデータセットを作成した。スクレイピングで収集した情報は、イラストとそのイラストのタイトル（例：石に刺さった剣を抜くアーサーのイラスト）である。Web サイト「いらすとや」[6]のイラストのタイトルには必ず末尾に「のイラスト」と記述されているため、類似度に影響を与える可能性がある。そこで、「のイラスト」の直前までをタイトル（例：石に刺さった剣を抜くアーサー）として保存する。また、Web サイト「いらすとや」[6]には様々なカテゴリが存在している。本研究では、カテゴリ「物語」を対象として 219 枚のイラストを収集した。

3 実験

本実験では、システムの評価を適切に行うため、イラストのある場面のみを対象とし、解析を行った。実験に使用するデータとして、青空文庫[15]で公開されている小説 20 冊を使用する。青空文庫[15]は、著作権が切れた文芸作品などを扱っていることが特徴として挙げられる。実験に使用した青空文庫[15]のテキストの一部を図 3 に示す。イラスト推薦の結果は推薦するイラストと元の挿絵を比較することと、挿絵を推薦する時に使用した文と推薦するイラストのタイトルのコサイン類似度を求めることで評価した。このとき、類似度を算出する前に、CNN[16]を使って画風変換を行うことでスタイルを統一した。元の挿絵との類似度が高ければ高いほど、元の挿絵の情報を多く抽出できているイラストであることを表す。青空文庫[15]のテキストには、[#@@@]を基本形とする [#3 字下げ] などのようにレイアウトやルビなどが含まれているため、削除した。また、松本ら[17]が、話し言葉は形態素解析の精度が下がると報告されていたため、会話文の記号である「と」を削除した。

```
[#挿絵] (fig51344_04.png, 横487×縦575) 入る]
むかしむかし あるところに 4ひきの こうさが おりました。
なまえは それぞれ
[# 5字下げ] フロブシー、
[# 4字下げ] モブシー、
[# 3字下げ] カンテル、
[# 2字下げ] ピーターです。
4ひきは おかあさんと いっしょに とってもおおきな モミのきの
したにある あなのなかに すんでいました。
```

図 3 テキストの一部。

3.1 場面分割の評価方法

TopicTiling[1]での場面一致度は再現率, 適合率, F 値によって評価する. TopicTiling[1]の既存手法である TextTiling[8]を提案した Hearst ら[8]の実験は, 被験者が判断した場面境界と TextTiling[8]の一致度を再現率, 適合率, F 値で評価を行っていた. そして, このときの F 値が 0.58 を超えていると, TextTiling[8]は十分に場면을分割できているという結果が得られていた. しかし, この方法は被験者の判断で結果が変わるため, 客観的に評価することが難しいと考えた. そのため, 本研究の評価は以下のように定義した式を使用し, 挿絵の挿入箇所と TopicTiling[1]の出力を比較する. また, 本実験で使用する TopicTiling[1]のトピックモデルは, トピックモデルの一般的な評価指標である Perplexity と Coherence で最良値が得られたモデルを用いる. 挿絵の挿入箇所とは, 青空文庫[15]のテキストにある [# 挿絵 n @.png 入る] (n: 文頭から数えた挿絵の番号, @: 任意のファイル名) を指す. 比較する手法として, 場面分割の既存手法である TextTiling[8]を用意した. TopicTiling[1]は文間で場면을分割する手法であるが, TextTiling[8]は単語間で場면을分割する手法である. そのままの TextTiling[8]を使用する場合, TopicTiling[1]と TextTiling[8]を比較することができないため, TextTiling[8]が文間で場면을分割できるように拡張した.

$$\text{再現率} = \frac{\text{場面境界との一致数}}{\text{挿絵の数}} \quad (2)$$

$$\text{適合率} = \frac{\text{場面境界との一致数}}{\text{TopicTiling[1]の場面境界の数}} \quad (3)$$

$$F\text{値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \quad (4)$$

3.2 重要文抽出の評価方法

重要文抽出の評価方法として一般的に ROUGE アルゴリズム[18]が使われている. ROUGE アルゴリズム[18]は ROUGE-N, ROUGE-L, ROUGE-S など複数の手法を持ち, 評価する重要文の集合と別の手法で用意した重要文の集合との類似度を求めて評価を行う. 今回の評価法は, N-gram (任意の N 文字が連続する文字列) の一致で評価する ROUGE-N と最長共通部分列の一致で評価する ROUGE-L を使用した. また, 重要文抽出に使用する実験データは Web サ

イト「小説家になろう」[19]の小説 100 冊である. 小説 100 冊は, 文の数が 1000 以上 5000 以下の条件を満たしているものをスクレイピングで収集した. 青空文庫[15]の小説で実験を行わなかった理由は, 青空文庫[15]にはテキストデータのみが用意されており, 小説の概要などの正解となる重要文が用意されていないからである. 一方, Web サイト「小説家になろう」[19]には, あらすじと本文が用意されている. しかし, このサイトの小説に挿絵は用意されていないため, 重要文抽出の実験のみ採用した. 本システムに採用した BERT[2]を用いる重要文抽出[3]は抽出型に分けられるため, 比較する手法は抽出型の既存手法であるグラフ構造のみで重要度を評価する LexRank[20]を用意した.

3.3 イラスト推薦の評価方法

小説の挿絵と解析結果に基づいて推薦したイラストの類似度を求めて評価を行う. 評価手法として, SSIM, MSE を使用した. このとき, 挿絵と本システムが出力するイラストのスタイルを統一するために画風変換を行う. また, 推薦した挿絵が文章解析後のテキストとどの程度類似しているのかを評価するためにコサイン類似度を使用した. 本研究は, オブジェクトが 1 つのイラストを推薦しているが, 青空文庫[15]に用意されている挿絵には複数のオブジェクトが存在している. そのため, 挿絵に複数のオブジェクトが含まれている場合は, OpenCV の findContours 関数で 1 つのオブジェクトを抽出した後に評価を行なった. また, Web サイト「いらすとや」[6]に用意されているイラストは単色の背景画像であるが, 青空文庫[15]の挿絵には複数の色が使用されている背景が存在する. 本実験では, 青空文庫[15]の挿絵の背景情報の影響を抑えるために, Canny 法[21]で輪廓を抽出して単色の画像に置き換えた.

4 実験結果

4.1 場面分割の実験結果

隣接する単語を含める範囲を決めるパラメータのウィンドウサイズを変更しながら TopicTiling[1]と TextTiling[8]を比較した結果を表 1 に示す. 表 1 より, TopicTiling[1]は既存手法の TextTiling[8]を上回る結果が得られた. また, 再現率, 適合率, F 値の最大値を得るウィンドウサイズは 10 であることが読み取れる. したがって, 場면을分割した後に行われる重要文抽出, 文章解析, イラスト推薦で扱うテキストは, ウィンドウサイズが 10 の TopicTiling[1]で場면을

分割したものである。TextTiling[8]が TopicTiling[1]に劣った理由を述べる。拡張した TextTiling[8]は文の類似度と閾値を比較して判定を行っていたが、各文の類似度がほぼ等しい場合適切に場面を分割することが難しくなる。各文の類似度がほぼ等しくなる理由として、文中に出現頻度が多い単語が周辺の単語に存在していれば、単語間の類似度に大きい変化がなくなり、文の類似度の平均にほぼ変化がないことが挙げられる。一方、TopicTiling[1]はトピックモデルを使用することで単語の出現に対して補正が行われ、類似度が大きく変化するため、適切に場面を分割することが可能になった。

表1 場面分割の実験結果.

手法(ウインドウ)	再現率	適合率	F 値
TopicTiling(5)	0.81	0.81	0.81
TopicTiling(10)	0.81	0.82	0.81
TopicTiling(15)	0.80	0.82	0.81
TextTiling(5)	0.46	0.77	0.50
TextTiling(10)	0.36	0.73	0.43
TextTiling(15)	0.40	0.72	0.45

4.2 重要文抽出の実験結果

既存手法のLexRank[20]と比較した実験結果を表2に示す。表2より、BERT[2]を用いる手法[3]がLexRank[20]を全て上回っている。すなわち、本システムの重要文抽出に採用した手法は、既存手法より有効であることが示される。LexRank[20]が劣った理由を述べる。LexRank[20]は、コサイン類似度でグラフを構築するため、各文の共通情報を持つ簡潔な文を抽出する可能性がある。この課題点が評価値を下げたと考えられる。本実験の結果は、尾崎ら[22]がニュース記事で実験を行なった結果に近い評価値が得られていることと、BERT[2]を用いる手法[3]は既存手法を上回ることは確認できた。本実験で使用したデータは小説のテキストであるため、全体の流れを考慮して重要文を判断する必要がある。しかし、本システムに使用している手法は、全体の流れを考慮して重要文を抽出するのではなく、学習済みモデルから重要文を予測して抽出する。解決する方法として、登場人物の関係や行動から場面の変化を高精度で予測できるようになることが挙げられる。

表2 「小説家になろう」[19]100冊の平均.

	ROUGE-1	ROUGE-2	ROUGE-L
BERT	0.33	0.06	0.18
LexRank	0.28	0.03	0.16

4.3 イラスト推薦の実験結果

実験結果を表3に示す。比較する分散表現としてDoc2Vec[12]を用意した。表3のSSIMより、BERT[2]とDoc2Vec[12]ともに評価値は0.72であった。このことから、青空文庫[15]の挿絵の情報を72%含んでいる。残りの28%の情報は、3つの問題点が影響していると考えられる。1つ目は、複雑なオブジェクトを含む挿絵から1つのオブジェクトを抽出する方法が、複数のオブジェクトを完全に削除できずに残った状態で出力されている可能性がある。その場合、適切に評価が行われなかったため評価値に影響を与える。2つ目は、イラスト内のオブジェクトの向きが考慮されていないことである。具体的には、挿絵のウサギが左を向いていることに対して、推薦しているウサギのイラストは正面を向いていると評価値が減少する。これを解決する方法として、オブジェクトの向きを統一して評価する方法が考えられる。この評価法の準備として、挿絵のオブジェクトの方向を自動的に判定する必要がある。しかし、イラスト内のオブジェクトの向きを自動的に判定する有効な方法は提案されていないため、今後の課題点となる。3つ目は、重要文抽出の影響を受けていると考えられる。4.2節の結果では、重要文抽出が有効に機能していないことが示されている。有効に機能しない場合、挿絵に必要な情報を抽出することは困難になる。そのため、挿絵に不必要な情報で推薦が行われた可能性がある。この課題点を解決するためには、小説のテキストを解析することで、場面の変化を読み取り、重要な文を適切に抽出する手法が必要になる。表3のMSEより、Doc2Vec[12]の方がBERT[2]より誤差が少ないことが読み取れる。すなわち、Doc2Vec[12]はBERT[2]より挿絵の色に近いイラストを推薦していると考えられる。最後に、コサイン類似度の実験結果について述べる。表3のコサイン類似度より、BERT[2]のコサイン類似度は0.79であるため、文章解析後のテキストの内容に対して高い関連度を持つ挿絵を推薦できていると考えられる。一方、Doc2Vec[12]のコサイン類似度は0.69であり、類似度は減少している。減少した理由として、Doc2Vec[12]の単語の内部状態が文脈によって変化しないことが挙げられる。すなわち、文脈を利用した推薦を行うにはBERT[2]の分散表現を用いることが有効である。

重要文抽出の影響を確認するために、重要文抽出を行わなかった結果を表4に示す。表3と表4より、イラスト同士を比較する評価値(SSIM, MSE)とコサイン類似度にはほぼ影響を与えていない。そのため、重要文抽出はイラスト推薦に対して十分に機能していないと考えられる。

表3 青空文庫[15]20冊の評価値の平均.

	SSIM	MSE	cosine
BERT	0.72	254.18	0.79
Doc2Vec	0.72	246.61	0.69

表4 重要文抽出を行わなかった時の結果.

	SSIM	MSE	cosine
BERT	0.72	254.28	0.78
Doc2Vec	0.72	247.66	0.69

5 おわりに

本研究では、場面分割、重要文抽出、文章解析、イラスト推薦の順で解析を行なった。BERT[2]で得られる分散表現に対してコサイン類似度を計算することで適切な挿絵を推薦するシステムを提案した。青空文庫[15]の小説20冊で実験を行なった結果、本システムは十分な精度で挿絵を推薦していることと、BERT[2]の分散表現を使用してコサイン類似度を計算することが有効に機能していることが示された。さらに精度を向上させるためには、重要文抽出を有効に機能させる必要がある。本実験では、BERT[2]の分散表現や Doc2Vec[12]の分散表現を用いて重要文を抽出していた。しかし、これらの分散表現は内容に最も関係している文を見つけることが困難であるという結果が得られた。人物の相関図などのテキスト情報以外の情報も機械に与えることで、この問題を解決できる可能性がある。そのため、テキスト以外の情報が重要文抽出にどの程度影響を与えるのかについて調査を進める。

参考文献

- [1] Martin Riedl, Chris Biemann, TopicTiling: A Text Segmentation Algorithm based on LDA, ACL 2012 Student Research Workshop, pp. 37-42, (2012)
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint arXiv:1810.04805, (2018)
- [3] Yang Liu, Mirella Lapata, Text Summarization with Pretrained Encoders, EMNLP-IJCNLP, pp. 3730-3740, (2019)
- [4] 工藤 拓, 松本 裕治, チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834-1842, (2002)
- [5] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto, Applying Conditional Random Fields to Japanese Morphological Analysis, EMNLP, pp. 230-237, (2004)
- [6] いらすとや, <https://www.irasutoya.com/>, 最終閲覧日: 2020年11月5日
- [7] 西岡 渉, 沼尾 雅之, 小説を対象とした知的ナビ機能の付いた電子書籍リーダーの提案, DEIM Forum 2016 E8-5, (2016)
- [8] Marti A. Hearst, Multi-Paragraph Segmentation of Expository Text, the 32nd Annual Meeting of Association for Computational Linguistics, pp. 9-16, (1994)
- [9] 村井 聡一, 牛尼 剛聡, 電子化された小説の選別を支援する「立ち読み」インタフェース, 情報処理学会論文誌, Vol. 6, No. 4, pp. 24-34, (2013)
- [10] Karen Spärck Jones, A Statistical Interpretation of Term Specificity and its Application in Retrieval, Journal of Documentation, Vol. 28, No. 1, pp. 11-21, (1972)
- [11] 小坂 直輝, 小林 哲則, 林 良彦, 隠れた良作を推薦可能な Web 小説レコメンドシステムの提案, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会(第23回)研究発表予稿集, pp. 1-8, (2019)
- [12] Quoc Le, T Mikolov, Distributed Representations of Sentences and Documents, International Conference on Machine Learning, pp. 1188-1196, (2014)
- [13] David M. Blei, Andrew Y. Ng, Michael I. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research 3, pp. 993-1022, (2003)
- [14] 京都大学黒橋・村脇研究室, <http://nlp.ist.i.kyoto-u.ac.jp/index.php>, 最終閲覧日: 2020年11月5日
- [15] 青空文庫, <https://www.aozora.gr.jp/>, 最終閲覧日: 2020年11月5日
- [16] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, IEEE, Vol. 86, No.11, pp. 2278-2324, (1998)
- [17] 松本 裕治, 伝 康晴, 話し言葉の形態素解析, 自然言語処理研究会報告, Vol. 143, pp. 49-54, (2001)
- [18] Chin-Yew Lin, ROUGE: A Package for Automatic Evaluation of Summaries, Workshop on Text Summarization Branches Out, Post Conference Workshop of ACL 2004, pp. 74-81, (2004)
- [19] 小説家になろう, <https://syosetu.com/>, 最終閲覧日: 2020年11月5日
- [20] G. Erkan, D. Radev, LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization, Journal of Artificial Intelligence Research 22, pp. 457-479, (2004)
- [21] John Canny, A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, pp. 679-698, (1986)
- [22] 尾崎 花奈, 小林 一郎, 事前学習モデルと潜在トピックを用いた文書要約への取り組み, 言語処理学会第26回年次大会発表論文集, p828-831, (2020)