

doc2vec による学術論文の被引用箇所推定の一手法

田邊 俊介[†] 太田 学^{††} 高須 淳宏^{†††} 安達 淳^{†††}

[†] 岡山大学工学部情報系学科 〒700-8530 岡山県岡山市北区津島中 3-1-1

^{††} 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

^{†††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†]pfe88ycw@s.okayama-u.ac.jp, ^{††}ohta@de.cs.okayama-u.ac.jp, ^{†††}{takasu, adachi}@nii.ac.jp

あらまし 学術論文では、著者が引用した手法やデータなどについて記載されている文献を参考文献として記載する。そのためこれらの参考文献は論文の内容を深く理解するための助けとなる。しかし、著者が参考文献のどの部分に参照しているかは記載されていないことが多い。そのため多数の参考文献の全てを確認すると労力がかかる。そこで本稿では、論文を閲覧するユーザの負担を減らすことを目的として、その参考文献中の適切な参照箇所を“被引用箇所”として推定する手法を提案する。まず、論文中で参考文献を参照している文章を“引用箇所”として抽出する。そして、引用箇所と参考文献内の文章を doc2vec によりベクトル化し、引用箇所の特徴ベクトルと最も類似した特徴ベクトルを持つ文を被引用箇所と推定する。提案手法を評価するために、英語論文を対象に被験者に推定した被引用箇所の適切性を問うアンケートを実施した。その結果、引用箇所の引用目的により被引用箇所として有効な文を推定することができることを確認した。

キーワード 引用意図, 閲覧支援, 被引用論文, doc2vec

1. はじめに

学術論文には多くの場合参考文献がある。その役割としては、著者が論文内で主張する内容の根拠を示す、先行研究との関係を位置づける、他の研究との比較により実験の結果を強調する、など多岐にわたる。いずれも論文内容を補完するものであり、これらを参照することで読者はより理解を深めながら論文を読み進めることができる。しかし、一つの論文に対して百件の参考文献が記載されている場合もあり、全てを閲覧するのは読み手にとって大きな負担となる。

そのため論文中の引用に対して読み手が求める情報を含む適切な文章を参考文献内で特定し提示する手法が提案されている。石井らは [1], 論文本文に含まれる引用を表す文字列を含む文を“引用箇所”, それに対する読み手に提示するのにふさわしい文章を“被引用箇所”とし、まず人手で収集した手がかり語を用いて引用箇所からその引用意図を推定した。そして、その引用意図をもとに被引用箇所が含まれると考えられる参考文献中の節を推定し、手がかり語を用いてその節の中で適切な被引用箇所を特定し提示するという方法を提案した。しかしこの方法では、被引用箇所の特定のために人手で集めた手がかり語を用いるため、様々な分野の論文への応用は困難であるという課題があった。そこで吉次らは、引用意図分類に自動抽出した手がかり語を用いる方法と doc2vec [3] よりベクトル化した引用文を分類する方法を提案した [2]。本研究では、doc2vec による文章の類似度を用いて被引用箇所を推定する方法を提案する。

以下に本稿の構成を示す。2 節で本研究に関連する研究を紹介し、3 節で被引用箇所の推定方法について述べる。4 節で推定した被引用箇所の適切性等に関する評価実験を示し、5 節で本研究をまとめ今後の課題について述べる。

2. 関連研究

2.1 論文閲覧支援システム

近年、学術論文は電子化し公開することが一般的である。そのためこれらの閲覧の利便性を向上させることを目的としたシステムとして様々なものが提案されている。

阿辺川ら [6] [7] は PDF 形式の論文を画像に変換し、Web ブラウザ上で閲覧する文献閲覧システム Sidenoter を提案した。このシステムでは、文書の拡大縮小表示、背景色変更などが可能な閲覧機能、全文検索のほかタイトル、著者、セッション名などの属性を指定した検索が可能な検索機能、表示している論文の補足情報をページの左右に表示する脚注表示機能を使用することができる。特に脚注表示機能では、Wikipedia をリソースとした見出し語の説明の表示や、TF-IDF によって重み付けされたベクトル空間モデルでの類似性を用いた関連論文の列挙など、本文と結び付けられた外部の関連情報を表示できる。

2.2 関連論文の分類

難波ら [8] [9] はサーベイ論文作成の支援を目的とした関連論文の分類に関する研究を行った。具体的には、照応詞や接続詞などを“cue word”とし、これとルールを用いて論文中の参照箇所を抽出して分析した。まず論文が他の論文を参照する理由として、新しい理論を提案する根拠として参照する“論説根拠型”, 関連研究と比較するために参照する“問題点指摘型”, 論説根拠型と問題点指摘型のどちらにも当てはまらない“その他型”の 3 種類の参照タイプを定義した。そして、問題点指摘型の参照関係に注目しこれを利用することで、関連論文を比較するサーベイ論文作成支援システムを作成した。さらに難波らは、被参照論文の参照理由を考慮することで論文間の類似度を算出し、学術論文を組織化する手法を提案した。具体的には、2 論文

間で同一論文を共に参照し、かつそれらの参照タイプが一致している参照箇所を数える事で、それらの類似度を算出した。

2.3 doc2vec

doc2vec [3] は、任意の長さの文章をベクトル化する Paragraph Vector の一つである。この手法は、文章中の単語を前後関係を元に学習し分散表現を得ることができる word2vec [4] [5] を拡張したものである。3層のニューラルネットワークを使用し、文中に出現する単語を予測するよう学習することで、その文章の特徴ベクトルを獲得する。学習モデルには、word2vec の CBOW を拡張した PV-DM と Skip-gram を拡張した PV-DBOW の2種類がある。doc2vec は、文書の感情極性判定や文書検索の評価実験において、長文、短文に関わらず、単語の語順を考慮しない Bag-of-Words などのそれまでの手法を大幅に上回ることが報告されている。

3. 被引用箇所の推定

3.1 概要

本稿で提案する被引用箇所推定の方法について、先行研究である [1] との相違点を含めて詳しく説明する。被引用箇所推定では最初に論文中の引用箇所を特定し、次にそれをもとに被引用箇所を探索する被引用論文の節を特定する。最後に引用箇所と被引用箇所探索節を用いて被引用箇所を推定する。まず、3.2節で被引用箇所を推定するための情報を得るための引用箇所の特定方法について述べる。そして、3.3節で引用意図の分類とその結果を用いた被引用箇所探索節の特定方法について述べる。3.4節で2種類の被引用箇所推定方法について説明した後、3.5節でユーザに対して提示する適切な被引用箇所の文章量について述べる。なお本稿では英語論文を対象とする。

3.2 引用箇所の特定

論文の PDF ファイルから引用箇所を特定する方法について説明する。まず、pdftotext [10] を用い、PDF ファイルをテキスト化する。そして、ParsCit [11] を用いてテキスト化した論文から本文のみを取り出す。ParsCit は CRF を用いて論文のテキストから書誌情報を抽出する。このプログラムのオプションの、テキストに論文構成要素タグを付与する extract.section コマンドを使用し、bodyText タグが付与された部分を論文本文とした。この論文本文には論文題目や図表、参考文献などは含まれない。そして、引用を示す “[1]” のような表現（以下、参考文献マーカ）を手がかりとして引用箇所を特定する。

石井らは、引用したい内容の手掛かりは参考文献マーカを含む1文に含まれると考え、この1文のみを引用箇所とした [1]。しかし、この定義では、その1文と深く関係するかもしれない前後の文を考慮していない。そこで、本研究では O’Conner の手法 [12] を参考に、以下の条件によって定まる文章を引用箇所とする。

まず、参考文献マーカを R 、 R を含む1文を S 、 S_{-1} を S の直前の1文、 S_1 を S の直後の1文とし、次の手順で引用箇所を選択する。

- (1) S を選択する。
- (2) S の先頭12語に接続を表す単語 (this, these, those,

such, same, similar, similarly, they, their, former, latter) (以下、コネクタ) があれば S_{-1} も選択する。

(3) S_{-1} の先頭7語にコネクタがあれば S_{-2} も選択する。

(4) S_{-2} の先頭4語にコネクタがあれば S_{-3} も選択する。

(5) S_1 の先頭7語にコネクタがあり、 R を含まなければ S_1 も選択する。

(6) S_1 が選択され、 S_2 の先頭4語にコネクタがあり R を含まなければ S_2 も選択する。

(7) S_{-1} と S が同じ段落にあり S_{-1} に他の R が含まれずどちらかが14語未満ならば S_{-1} も選択する。

(8) S と S_1 が同じ段落にあり S_1 に他の R が含まれずどちらかが14語未満ならば S_1 も選択する。

このようにして選択した文章を引用箇所とすることで、1文のみと比べて被引用箇所を推定するための手がかりが増える。

3.3 引用意図の分類と被引用箇所探索節の特定

[1] で提案された引用意図のクラスとその分類方法、それらを用いた被引用箇所探索節の特定方法について説明する。まず3.3.1項で引用意図の分類クラスについて説明する。そして、3.3.2項で引用意図分類の手順を述べ、3.3.3項で被引用箇所探索節の特定について述べる。

3.3.1 引用意図の分類クラス

[1] では、NTCIR-9 [13] に投稿された論文の引用箇所を手で分析し、引用意図の分類クラスを “Group”, “Method”, “Result”, “Data”, “Equation”, “Other” の6クラスと定めた。各引用意図クラスについて以下にまとめる。

- Group

タスクやフォーラム、ワークショップの詳細が特に引用したい内容であるもの。

- Method

研究に使用する、または著者の手法と比較する既存の手法や考えが特に引用したい内容であるもの。

- Result

既存研究の実験結果が特に引用したい内容であるもの。

- Data

実験等のデータが特に引用したい内容であるもの。

- Equation

計算式の詳細が特に引用したい内容であるもの。

- Other

上記5つの、どのクラスにも当てはまらないもの。

3.3.2 引用意図の分類方法

[1] では、NTCIR-9 [13] に投稿された論文の引用箇所を手で分析し、各分類クラスの手がかり語を収集した。この手がかり語を「CI (Citation Intention) 手がかり語」とした。CI 手がかり語の例を表1に示す。そして、引用箇所の文の単語と CI 手がかり語を1単語ずつ比較し、完全に一致する CI 手がかり語の数を引用意図ごとに数え、以下のように分類した。

(1) Equation の CI 手がかり語の数が1以上あれば “Equation” とする。

(2) (1) の条件を満たさず、かつ “Data” の CI 手がかり語が1以上あれば “Data” の意図とする。

表 1 CI 手掛かり語の例 [1]

引用意図	CI 手掛かり語
Group	work, workshop, task, campaign
Method	system, base, use, develop, toolkit, improve, method, model
Result	analysis, performance, participant, achieve
Data	collection, library, set
Equation	calculate, formula

表 2 被引用論文の被引用箇所を探索する節の種類 [1]

引用意図	被引用箇所を探索する節の種類
Group	introduction
Method	method
Result	method, evaluation, conclusions
Data	introduction, method, evaluation, conclusions
Equation	method

(3) (1) と (2) の条件を満たさず、かついずれかの CI 手掛かり語が 1 以上あれば、一番 CI 手掛かり語の数が多い引用意図とする。

(4) (1) から (3) の全ての条件を満たさない場合“Other”に分類する。

3.3.3 被引用箇所探索節の特定

[1] では、ParsCit [11] の持つオプションコマンド `extract.section` で付与されるラベルから、論文の一般的なセクション名を示す“genericHeader”を利用し、本文の節を“keywords”, “abstract”, “introduction”, “method”, “evaluation”, “conclusions”, “acknowledgments”の 7 種類に分類した。そして、表 2 に示した対応表をもとに、引用意図ごとに被引用箇所を探索する節を特定した。

3.4 被引用箇所の推定

[1] では、引用箇所とそれに対応する被引用箇所には同じ名詞が含まれている可能性が高いとの分析結果から、被引用箇所を分析して得た「CP (Cited Parts) 手掛かり語」と引用箇所に出現する単語を用いて被引用箇所を推定した。本稿でも、引用箇所と被引用箇所にはそれぞれに同一の固有名詞や表現が含まれており、類似した文章であると仮定する。しかし、doc2vec により引用箇所と被引用論文中の各文のベクトルを算出し、ベクトル間の類似度によって被引用箇所を特定する方法を提案する。なぜなら、被引用箇所の特定のために人手で集めた手がかり語を用いないため、様々な分野の論文へ応用しやすいと考えたからである。

3.4.1 [1] の手がかり語による被引用箇所の推定

比較のため [1] での被引用箇所の特定方法について述べる。[1] では、引用箇所を TreeTagger [14] で形態素解析して得られた形容詞、名詞と、分析の結果得られた被引用箇所の特定のための手がかり語「CP (Cited Parts) 手掛かり語」の 2 種類の手がかり語を使用して被引用箇所を特定した。CP 手がかり語の例を表 3 に示す。具体的には、3.3.3 項の方法で特定した節中の各文について文中に含まれる 2 種類の手がかり語の数を数え、その文を構成する単語数で割る。そして、この値が最も高かった

表 3 CP 手掛かり語の例 [1]

引用意図	CP 手掛かり語
Group	organizer, workshop, focus, aim
Method	form, provide, include, infer
Result	suggest, experiment, evidence
Data	choose, expand
Equation	function, :, equation, calculate

1 文を被引用箇所とする。

3.4.2 doc2vec による被引用箇所の推定

3.2 節の方法で特定した引用箇所と 3.3 節の方法で特定した被引用箇所探索節の各文の特徴ベクトルを doc2vec [3] を用いて生成する。本研究では、被引用論文の文章を 1 文ずつに分け、それぞれの特徴ベクトルを 100 次元、窓サイズを 4、学習モデルを PV-DM、イタレーションを 50 とし、各文の特徴ベクトルを生成した。

doc2vec により生成された各文のベクトルの類似度を算出することで、類似した文を見つけることができる。そこで、引用箇所の文章のベクトルと最も類似度が高いベクトルをもつ 1 文を適切な被引用箇所と特定する。本研究では比較のため引用箇所として、3.2 節の方法で特定した文章に加え、[1] の手法で用いた 1 文のみ場合も実験で検証した。

3.5 被引用箇所の文章量

ユーザに提示する被引用箇所の文章量として [1] では、特定された 1 文、その 1 文の前後の文を合わせた 3 文、あるいはその 1 文を含む 1 段落を実験的に採用した。しかし、1 文の場合では説明が少なく、一方 1 段落の場合には不必要な情報も多くなる、などの問題があった。そこで本稿の実験では、被引用箇所を 3 文または、3.2 節の O’Conner の手法に従って特定された 1 文に前後の文を加えた文章の 2 種類として比較する。

4. 評価実験

まず、予備的な実験として、3.2 節で述べた引用箇所の特定方法により引用箇所が 1 箇所あたり何文になるかを確かめた。次に、3.4 節で述べた手法で特定した被引用箇所の適切性の評価をするため、アンケートによる被験者実験を行った。

4.1 引用箇所の文章量

3.2 節の方法で引用箇所が何文となるかを確かめた。対象は NTCIR-9 [13] に投稿された論文中の引用箇所 10 件である。結果は 1 文のみが 3 箇所、2 文が 5 箇所、3 文が 1 箇所、4 文が 1 箇所となり平均 2.0 文となった。

4.2 被引用箇所の適切性の評価

3.4 節で述べた doc2vec を用いた被引用箇所推定方法の適切性を調べるため、提案手法で得た文章が引用箇所の説明として適切であるかをアンケートにより被験者に評価させた。また、同時に [1] の手法で得た文章と、被引用論文の概要から人手で抜き出した文章も評価させた。さらに、各引用箇所に対して提示されるこれらの文章の中から、最も適切だと思う文章を選択させた。

4.2.1 実験内容

本実験では、岡山大学大学院自然科学研究科の大学院生2名と岡山大学工学部情報系学科の学部4年生3名の計5名に対してアンケートを行った。アンケートの内容は、ある1つの引用箇所に対して被引用箇所の候補を複数提示し、それぞれについて「提示する引用箇所の補足や説明として適切であるか」という質問を行い、「適切である」、「やや当てはまる」、「まったく当てはまらない」の三択で答えさせるというものである。また、1つの引用箇所につき1つだけ、説明として最も適切だと思う文章を提示されている文章の中から選択させる。NTCIR-9 [13] に投稿された論文中で正しく引用意図が分類でき、被引用論文内の被引用箇所探索節を特定できた10箇所の引用箇所についてアンケートを作成して評価した。設問に用いる引用箇所は、“Group”, “Method”, “Result”, “Data”, “Equation”の各クラスの引用箇所を2箇所ずつ選んだ。提示する被引用箇所の候補は以下の10種類である。

- A: [1]の手法で推定した1文に前後の文を1文ずつ加えた3文
- B: 被引用論文の概要から被引用箇所に適すると人手で判断し抜き出した3文
- C: 引用箇所を1文とし、被引用論文の被引用箇所探索節中から特定した1文に、3.2節のO'Connerの手法に従って前後の文を加えた文章
- D: 引用箇所を1文とし、被引用論文の被引用箇所探索節中から特定した1文に前後の文を1文ずつ加えた3文
- E: 引用箇所を3.2節のO'Connerの手法に従って得た文章とし、被引用論文の被引用箇所探索節中から特定した1文に3.2節のO'Connerの手法に従って前後の文を加えた文章
- F: 引用箇所を3.2節のO'Connerの手法に従って得た文章とし、被引用論文の被引用箇所探索節中から特定した1文に前後の文を1文ずつ加えた3文
- G: 引用箇所を1文とし、被引用論文の本文全文から特定した1文に3.2節のO'Connerの手法に従って前後の文を加えた文章
- H: 引用箇所を1文とし、被引用論文の本文全文から特定した1文に前後の文を1文ずつ加えた3文
- I: 引用箇所を3.2節のO'Connerの手法に従って得た文章とし、被引用論文の本文全文から特定した1文に3.2節のO'Connerの手法に従って前後の文を加えた文章
- J: 引用箇所を3.2節のO'Connerの手法に従って得た文章とし、被引用論文の本文全文から特定した1文に前後の文を1文ずつ加えた3文

CからJはそれぞれ本稿の提案手法で条件を変えて得た文章であり、まとめると表4となる。引用箇所により推定される被引用箇所が重複する場合があるため、実際に提示した被引用箇所の候補は1つの引用箇所につき3から10となった。また、被験者が引用箇所の内容を理解するため、その引用箇所を含む段落と論文の概要も同時に提示した。被引用箇所の候補の提示順は無作為に変えた。

表4 本稿の手法で提示した文章を得るための条件

記号	C	D	E	F
被引用箇所文章量	O'Conner	3文	O'Conner	3文
引用箇所文章量	1文		O'Conner	
被引用箇所探索範囲	被引用箇所探索節			
記号	G	H	I	J
被引用箇所文章量	O'Conner	3文	O'Conner	3文
引用箇所文章量	1文		O'Conner	
被引用箇所探索範囲	全文			

4.2.2 考察

4.2.1項で実施した被験者実験の結果をグラフにまとめる。各引用意図の引用箇所に対する被引用箇所の適切性についての被験者の回答の割合は図1～図5となり、全てをまとめると図6となった。また、各引用意図の引用箇所に対して、それぞれの被引用箇所として最も適切であると選択された被引用箇所の割合は図7となった。

図1, 図2では引用意図“Data”, “Group”の設問において、[1]の手法で得た文章は本稿の手法で得た文章より適切性が高いことがわかる。本稿の手法で得た文章は、特に引用意図“Data”の設問では、図1に示すように引用箇所と関係ないものばかりであった。よって、“Data”や“Group”などが引用したい内容である場合、引用箇所に含まれている名詞や形容詞などを手がかりとして被引用箇所を特定するのが有効であることがわかる。引用意図“Equation”の設問の結果を示す図3では、全ての被引用箇所候補の適切性が低いことがわかる。被引用論文の概要から人手で抜き出した文章Bは全体的に評価が高かったが、この引用意図に関しては低い評価であった。これは、本研究では被引用箇所を数式を除く本文のテキストの中から探索しているからであると考えられる。被引用論文のテキスト化の処理の際、数式、図、グラフを除いてしまうため、それらの情報は本研究の実験では提示できない。しかし引用意図“Equation”は、数式の詳細が特に引用したい内容であるため、テキストだけで説明するのは適切ではない。学術論文では本文で必ず一度は論文中で使用した数式、図、グラフを参照するため、これを利用して適切な情報を探索し特定するという方法が解決策になる可能性がある。図4では引用意図“Result”の設問においては、全ての被引用箇所候補の適切性が他の引用意図の設問に対して比較的高く、差がほぼ無いことがわかる。図5では引用意図“Method”の設問において、提案手法によって推定した被引用箇所の多くが引用箇所の説明として適切であると判断されている。特に、提案手法により推定した被引用箇所の中では、被引用論文の本文全文から被引用箇所を探索して得た文章の評価が高かった。図6から、全体的には、[1]の手法で得た文章の適切性が高く評価されていた。よって、被引用箇所推定における単語の一致率の重要性がわかる。また、概要から人手で抜き出した文章の被引用箇所としての適切性は、[1]の手法で得た文章とほぼ同等である。概要は、著者自身がその論文の要点をまとめた文章であり、そのなかから人手で選択しているので比較的高い評価を得たと考えられる。

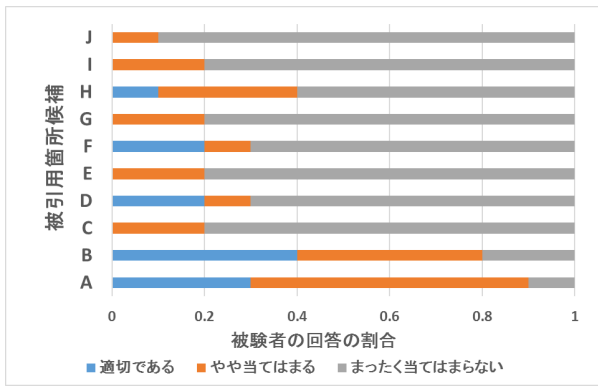


図1 被引用箇所の適切性評価 (引用意図: Data)

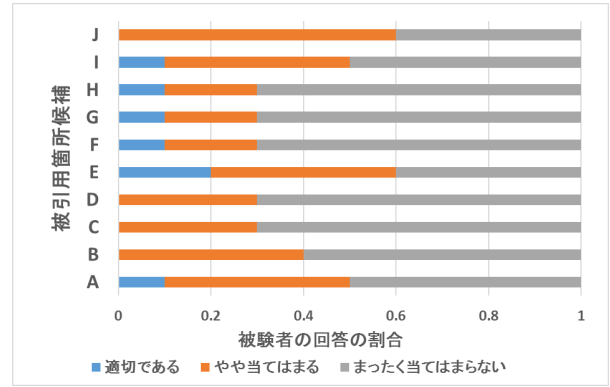


図3 被引用箇所の適切性評価 (引用意図: Equation)

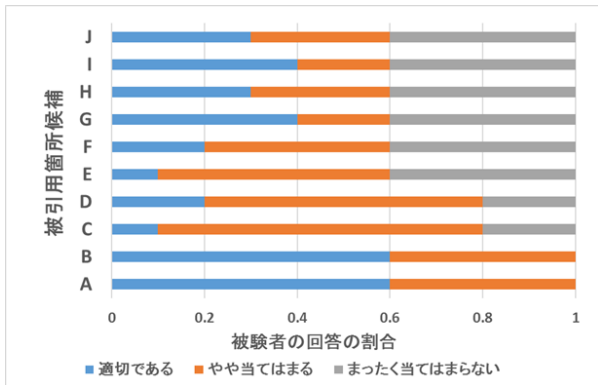


図2 被引用箇所の適切性評価 (引用意図: Group)

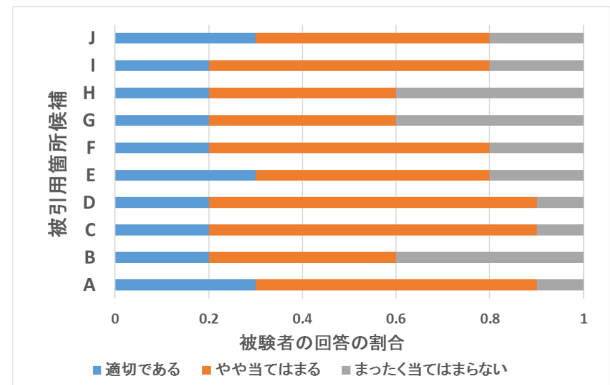


図4 被引用箇所の適切性評価 (引用意図: Result)

図6に示したように、本稿の提案手法では、推定されたCからJの文章のうち、被引用箇所探索時に被引用論文の節を特定して探索するより、全文から探索して推定されたGからJの文章の方が少し適切性が高かった。また、提示する被引用箇所としては、3.2節のO'Connorの手法で周辺の文を加えたものより、前後の分を1文ずつ加えた3文のほうが若干適切だったといえる。前者は、3.3.3項の手法で特定した被引用箇所探索節以外の節にも、被引用箇所として適切な文章を含む節がある可能性を示している。後者は、単純に3文という文章量が量として適当だったことが考えられる。3.2節のO'Connorの手法で周辺の文を加えた場合、1文のみの提示となる場合も多く、説明の文章量としては不十分であった可能性が高い。図7に示すように、全体的に概要から人手で抜き出した文章が最も適切な文章として選ばれることが多かったが、引用箇所の引用意図によっては本稿の提案手法で得られた文章Jがそれに並ぶ、またはそれより高い評価を得ることがあった。

5. まとめ

本研究では、論文内で引用される参考文献中の適切な参照箇所(被引用箇所)をdoc2vecにより推定する手法を提案した。提案した手法ではまず、論文の引用箇所を“[1]”のような表現を元に特定する。そして特定された引用箇所1文を手がかり語などを用いて引用意図クラスに分類した後、そのクラスを元に被引用論文から被引用箇所を探索する節を特定する。特定した引用箇所に関係の深い周辺の文を加えた文章と被引用箇所探索節の各文のベクトルをdoc2vecを用いて生成し、引用箇所

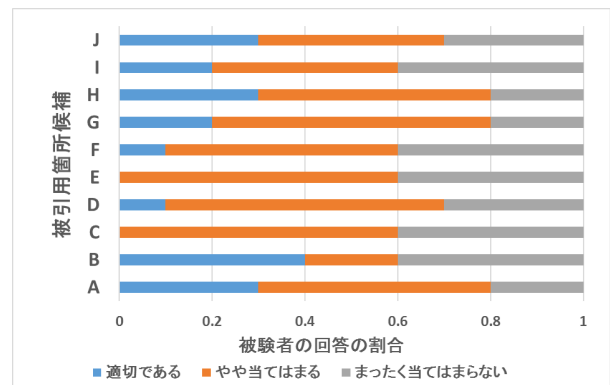


図5 被引用箇所の適切性評価 (引用意図: Method)

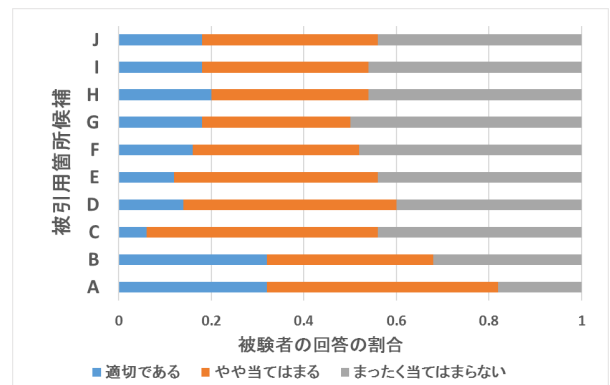


図6 全ての引用箇所に対する被引用箇所の適切性評価

の文章のベクトルと類似度の高い特徴ベクトルを持つ文に、関係の深い周辺の文を加えた文章を被引用箇所と推定した。

実験では、被験者に対して、被引用論文の概要から人手で抜

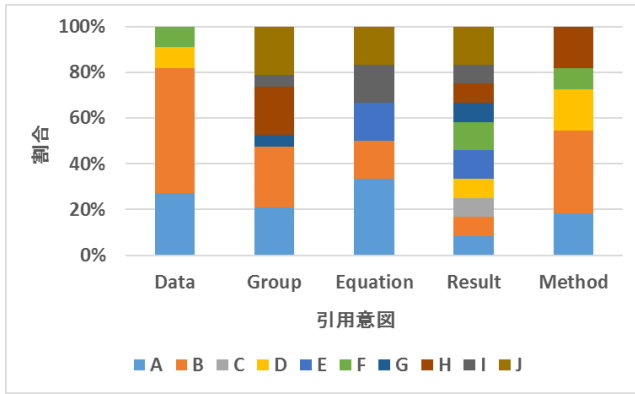


図7 被験者が選択した最も適切な被引用箇所

き出した文章、先行研究の手法で得た文章、本稿の提案手法で得た文章を提示し、それぞれが引用箇所の説明として適しているかを問うアンケートを実施した。結果として、著者自身が論文の重要部分をまとめた概要から人手で抜き出した文章と、固有名詞と形容詞や手がかり語の一致率を用いて被引用箇所を推定する先行研究の手法で特定した文章が、全体的には被引用箇所として高い評価を得た。一方、doc2vecを用いた本稿の提案手法で推定した文章も、引用箇所の引用意図によって有効な場合があることを確認した。

今後の課題として、引用箇所中に出現する固有名詞を加味した被引用箇所の探索方法の検討や、引用箇所の説明として、図、表、数式を加えて提示できる論文閲覧支援システムの検討などが挙げられる。

謝 辞

本研究の一部は、国立情報学研究所公募型共同研究の援助による。ここに記して深謝する。

文 献

- [1] 石井仁子, 太田学, 高須淳宏, “引用意図を利用した学術論文閲覧支援のための適切な被引用箇所の特定情報処理学会研究報告”, 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), F3-5, 2015.
- [2] 吉次 優, 太田学, 高須淳宏, “機械学習による学術論文の引用意図分類一手法”, IEICE technical report, Vol. 116, No. 340, pp. 45-69, 2016.
- [3] Le, Q. and Mikolov, T., “Distributed Representations of Sentences and Documents”, CoRR, abs/1405.4053, pp. 1-9, 2014.
- [4] Mikolov, T., Sutskever, I., Chen, K., Corrado G. and Dean, J., “Distributed representations of words and phrases and their compositionality”, Advances in Neural Information Processing Systems, pp. 3111-3119, 2013.
- [5] Mikolov, T., Chen, K., Corrado G. and Dean, J., “Efficient estimation of word representations in vector space”, arXiv preprint arXiv : 1301.3781, pp. 1-12, 2013.
- [6] 阿辺川武, 相澤彰子, “脚注表示機能を備えた論文閲覧システム

Sidenoter”, 言語処理学会, 第20回年次大会発表論文集, pp. 796-799, 2014.

- [7] 阿辺川武, 相澤彰子, “内部構造解析機能と脚注表示機能を備えた論文閲覧システム”, 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第7回研究会, pp. 13-18, 2014.
- [8] 難波英嗣, 奥村学, “論文間の参照情報を考慮したサーベイ論文作成支援システムの開発”, 自然言語処理, Vol. 6, No. 5, pp. 43-62, 1999.
- [9] 難波英嗣, 神門典子, 奥村学, “論文間の参照情報を考慮した関連論文の組織化”, 情報処理学会論文誌, Vol. 42, No. 11, pp. 2640-2649, 2001.
- [10] pdftotext, <http://www.foolabs.com/xpdf>
- [11] ParsCit, <http://aye.comp.nus.edu.sg/parsCit>
- [12] O’Conner, J., “Citing statements: Computer recognition and use to improve retrieval”, Information Processing & Management, Vol. 18, No. 3, pp. 125-131, 1982.
- [13] NTCIR-9 (The 9th NII Test Collection for IR Systems), <http://research.nii.ac.jp/ntcir/ntcir-9>
- [14] TreeTagger, <http://www.cis.uni-muenchen.de/schmid/tools/>