

共引用テキストを利用した論文間の関係抽出

正元 修平[†] 清水 敏之[†] 難波 英嗣^{††} 吉川 正俊[†]

[†] 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都府京都市左京区吉田本町

^{††} 広島市立大学大学院情報科学研究科知能工学専攻 〒731-3194 広島市安佐南区大塚東 3-4-1

E-mail: †shogen@db.soc.i.kyoto-u.ac.jp, †{shimizu,yoshikawa}@i.kyoto-u.ac.jp, ††nanba@hiroshima-cu.ac.jp

あらまし 研究者が論文をサーベイする際、二つの論文間の関係を提示することは、それらの位置づけや違いを理解するために重要である。本稿では、二論文を共引用している論文における引用部周辺テキストを“共引用テキスト”と呼び、共引用テキストの特徴と、これを用いて論文間の関係を抽出する手法について考察する。既存研究では、二論文間の直接の参照関係や、引用部周辺テキストを用いた関係提示については行われているが、共引用テキストを用いた関係提示は行われていない。共引用テキストを利用することで、多様な側面から論文間の関係を抽出できる可能性がある。

キーワード 共引用テキスト, 関係抽出, 共引用分析

1. はじめに

Web 関連技術の発展により、数多くの学術論文を電子的に閲覧・利用することが可能になった。研究者は、Google Scholar^(注1)や、Microsoft Academic Search^(注2)に代表される、学術論文の検索に特化した検索エンジンを用いることで、Web上で論文を検索し、自らの研究に関連する論文を閲覧して知見を得たり、必要に応じて、論文執筆の際にその論文を引用したりすることができる(以下、本稿ではこれを「論文のサーベイ」と呼ぶ)。特に、ある研究分野から別の研究分野に研究対象を変えたなどの理由で分野の知識に乏しい研究者や、そもそも研究を始めてからの日が浅く、経験が少ない研究者にとって、論文のサーベイは、研究分野について知る上で重要になる。しかしながら、利用可能な論文の数に比べて研究者の処理能力には限界があり、自らの研究に関連するすべての論文を閲覧することは非常に困難である。そこで、研究者による論文のサーベイを省力化し、支援するための技術が必要となってくる。

論文のサーベイにおける問題点は多数あるが、我々は、その中でも、「関連する二つの論文間の関係理解が困難である」という問題に着目する。ある論文 A と、別の論文 B が互いに関連を持っているとすると、それらの関係を提示することは、サーベイを行う研究者が二論文の位置づけや違いを理解するために重要である。例えば、PageRank アルゴリズムに関する初出の論文である Page らの論文(以下、本稿では「PageRank の起源論文」と呼ぶ)[1] と、PageRank アルゴリズムを発展させることで考案された Topic-Sensitive PageRank アルゴリズムに関する初出の論文である Haveliwala らの論文(以下、本稿では「Topic-Sensitive PageRank の起源論文」と呼ぶ)[2] の二論文をみると、これらは互いに関連を持っていると言える。しかし、これらの論文を始めて知った研究者は、そのままでは、

それぞれの論文がどのような違いを持っているか、また、研究分野においてどのように位置づけられているかを知ることができず、結果として、どちらの論文から読み進めるべきかがわからない可能性がある。そこで、例えば「PageRank の起源論文と、Topic-Sensitive PageRank の起源論文には、『後者は前者を発展させたものである』という関係がある」というような知識が得られれば、研究者は PageRank の起源論文の有用性と研究分野における位置づけが理解でき、その後のサーベイに生かすことができるだろう。本研究では、この問題を解決するために、二論文間の関係を抽出する手法を考える。

二論文間の関係を抽出するための既存手法として、論文間の直接の参照関係を用いる手法や、引用部周辺テキストを用いた手法がある。二論文間の直接の参照関係を用いる手法では、論文間の引用・被引用関係を用いて類似度を計算したり、これを可視化することで、研究者に論文の関係理解を促す。しかし、この手法は、論文間の参照関係のみを考慮し、論文の内容を全く考慮しないため、関係抽出という観点からは限界があると考えられる。それとは別に、「引用部周辺テキスト」と呼ばれる、ある論文を引用している論文が、その論文について言及している箇所の周辺テキストを取得し、関係抽出に利用している研究もある。引用部周辺テキストには、引用している論文の著者が、引用される論文をどのような観点で見ているかが記述されることが多いため、二論文間の関係抽出に用いることができる。しかし、この手法では、論文が直接的な引用・被引用関係にあることが必要になり、このような関係にない二論文については手法の適用ができないという問題点がある。

本研究では、まず「共引用テキスト」という概念を定義し、共引用テキストの分類を行うことで、二論文間の関係を抽出する手法について考察する。二論文の「共引用テキスト」とは、二論文を共引用している論文における引用部周辺テキストのことである。共引用テキストを用いることで、直接の参照関係にない二論文についても、それらの関係を抽出することができる。また、共引用テキストには、論文を引用している著者から見た、

(注1): <http://scholar.google.com/>

(注2): <http://academic.research.microsoft.com/>

それぞれの論文の違いや、研究分野における位置づけが明示されていることが多く、そのような記述を抽出することで、論文間の関係を多様な側面から理解することができる可能性がある。

本稿の構成は以下の通りである。2章では、本研究に関連する既存研究について述べる。3章では、本研究が対象とする問題を述べ、その問題に対してどのような手法を取ればよいかについて考察する。4章では、今後の研究に向けて行った予備調査の内容とその結果について述べる。5章では、本稿のまとめと今後の課題について述べる。

2. 関連研究

2.1 論文のサーベイ支援

1章でも述べたように、研究者にとって論文のサーベイは重要だが、この作業は大きな負担を伴う。論文データベース中の論文を構造化したり、論文の要約を提示したりすることで、論文のサーベイを支援する研究が行われている。Hoangら[3]は、論文の関連研究の項を、その論文が引用している論文の要約とみなし、これらを部分に分けてサブピックに分類している。また、Qazvinianら[4]は、論文が複数の技術的貢献を含むことに着目し、対象となる論文を引用している論文における記述の集合をクラスタリングしてからその代表値を抽出することで、それぞれの技術的貢献に対応する要約文を生成している。他にも、論文データベースから研究トピックを抽出する研究[5]や、論文のトピックの変遷を解析する研究[6][7][8][9]が行われている。

2.2 論文の引用解析

関連する二論文間の関係を抽出するための手法として、二論文間の直接の参照関係を利用する手法が挙げられる。論文間の引用を解析する手法として、書誌結合[10]、共引用分析[11]などが知られている。書誌結合とは、二論文がどれだけ同じ論文を引用しているかをを用いて、論文の類似度を測る手法である。この手法は、「同じ論文を引用している論文は類似した主題の論文である」という仮定に基づいている。また、共引用分析とは、二論文がどれだけ同じ論文に引用されているか（共引用されているか）を用いて、論文の類似度を測る手法である。この手法は、「同じ論文に引用されている論文は類似した主題の論文である」という仮定に基づいている。いずれの手法も、論文の引用関係をグラフ的な手法で解析することで、類似度を計算している。共引用分析を発展させた手法として、江藤[12][13]は、従来の共引用分析が論文中の文脈を考慮していないことを問題点とし、共引用における「引用箇所の間隔」を定義している。この研究では、「引用箇所の間隔」に基づいて、共引用を、非同一段落における共引用、同一段落における共引用、同一文における共引用、同一箇所における列挙共引用の4種類に分類し、共引用された二論文の類似度がこの順で大きくなる傾向にあることを実験により示している。

2.3 引用部周辺テキスト

研究者は、自らの論文中で他の論文を引用する際、「 \times 」³⁾は \times する手法を提案している。」「 \times 」³⁾についての実験では、 \times という結果が出ている⁵⁻⁷⁾」などといったように、引用し

たい論文についての言及を論文本文中で行う。このような、ある論文を引用している論文が、その論文について言及している箇所の周辺テキストのことを、「引用部周辺テキスト」と呼ぶ。ある論文の引用部周辺テキストには、その論文と、その論文が引用している論文との関係を示すような記述があることが多く、論文間の関係抽出には非常に重要な要素であると考えられる。難波ら[14][15]は、2.2節で述べた書誌結合を改良し、論文における個々の引用を、論文内の記述の根拠となる引用である論説根拠型、引用した論文の問題点を指摘する引用である問題点指摘型、論説根拠型にも問題点指摘型にも当てはまらない引用であるその他型の3つの「参照タイプ」に分類している。その上で、彼らは、参照タイプを用いた論文の類似点・相違点の抽出や、論文の組織化などを行っている。同様の研究として、Teufelら[16]は、引用の種類を12種類に分類し、手がかり語などの素性を用いた機械学習を行っている。本研究は、直接の参照関係のない論文間の関係についても抽出できるという点が、これらの研究と異なる。

2.4 論文間の関係抽出

本研究では、二論文間の関係抽出を行う。論文間の関係を抽出する研究として、Nanbaら[17]は、論文間の参照関係に加え、2.3節で述べた「参照タイプ」や、引用部周辺テキストを可視化してユーザに提示するツールを提案している。また、Heら[18]は、論文本文の情報と論文間の引用情報を解析し、そこから研究トピックの時系列的な発展過程を抽出・可視化している。研究トピックの発展過程は、研究トピックを節点、研究トピック間の発展関係を有向枝としたグラフで表され、研究トピック間の発展関係には、「同じトピックである」、「似たトピックである」、「新出のトピックである」、「全く関係のないトピックである」の4種類のラベルが貼られる。これらとは別に、吉田ら[19]は、論文間の参照関係を用いて論文のクラスタリングを行い、研究のマクロな流れを抽出する手法を提案している。この際、引用・被引用の関係にある二論文について、これらの論文を共引用している論文から「論文Aを引用しているならば論文Bを引用している」といったようなアソシエーションルールを抽出し、この情報を用いてクラスタリングを行っている。この手法は、二論文間の類似度を求めることしかできない従来の共引用分析とは異なり、二論文間の関係の方向性を考慮することができ、研究の流れ抽出に利用できるという利点がある。この研究は、論文間の参照関係のみを用いたものであり、論文の内容は考慮していない点が、本研究と異なる。

3. 共引用テキストからの関係抽出

本研究では、「共引用テキスト」を用いることで、二論文間の関係を抽出するための手法について考察を行う。本章では、「共引用テキスト」の定義とその抽出方法、また、そこからどのように二論文間の関係抽出を行うかについての考え方を述べる。

3.1 共引用テキスト

2.3節でも述べたように、引用部周辺テキストとは、ある論文を引用している論文が、その論文について言及している箇所の周辺テキストのことである。引用部周辺テキストの定義はい

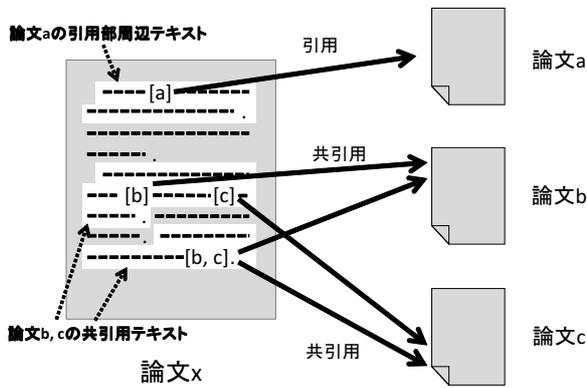


図1 引用部周辺テキストと共引用テキスト

くつか考えられるが、本稿では、簡単のために、引用部が含まれる一文を引用部周辺テキストと定義する。ここで、ある論文が1つの論文を複数回引用することが起こり得るが、引用部周辺テキストは、毎回の引用ごとに定義されるとする(すなわち、1つの参考文献に対して、複数の引用部周辺テキストが存在し得る)。例として、図1は、論文xと論文a, b, cの引用関係と、論文xにおけるこれらの論文の引用部を図示したものである。この図における論文xの、1段落目の1文目、2段落目の1文目、2段落目の3文目は、それぞれ、論文xにおける、論文a、論文bと論文c、論文bと論文cの引用部周辺テキストに相当する。

2.2節で述べたように、共引用における引用部間の間隔は、共引用されている論文の類似度に影響するとされている[12][13]。我々は、その中でも比較的類似度が高いとされる、同一文における共引用だが、同一箇所における列挙共引用ではない共引用(区別のため、以下、本稿ではこれを「同一文における共引用」と呼ぶ)と、同一箇所における列挙共引用に着目する。論文Xにおいて、ある論文Aと別の論文Bが、同一文中で共引用されているか、同一箇所で行挙共引用されている場合、その文を「論文Xにおける論文Aと論文Bの共引用テキスト」と定義する。図1の場合、2段落目の1文目、2段落目の3文目は、論文xにおける論文b, cの共引用テキストとなる。共引用テキストは、共引用されている各々の論文についての引用部周辺テキストとみることもできる。

3.2 共引用テキストの分類

共引用テキストは、これを論文の中から抽出してそのまま提示するだけでも、研究者が二論文の関係を知るには十分な情報を含んでいると考えられる。しかしながら、数多く共引用されている論文の共引用テキストをそのまますべて提示すると、情報量が非常に大きくなり、研究者が扱うことのできる量を超えてしまうと考えられる。そこで、我々は、共引用テキストから、二論文の関係を知るのに適切な情報のみを抽出し、これを研究者に提示することを考える。

共引用テキストから二論文間の関係を抽出する手法としては、共引用テキストを自然言語処理的な手法を用いて解析することで、関係に相当する単語や表現を取得する手法が考えられる。しかし、電子的に利用できる論文の数は膨大であり、自然言語

処理的な手法では、妥当な計算時間で関係を抽出できない可能性がある。そこで、本研究では、共引用テキストに対して、それがどのような関係を包含しているかという観点から分類を行う。その上で、共引用テキストを各々のタイプに機械的に分類する手法を提案することにより、二論文間の関係を抽出することとする。

以下に、我々が提案する2種類の関係の分類を示す。

(1) 共通型

共引用テキストが共通型であるとは、共引用テキストによって共引用されている各々の論文が、文中で並列の関係にある場合のことを指す。共通型の共引用テキストは、さらに2種類に分類することができる。それらの分類を以下に示す。

● 共通部分型

共引用テキストが共通部分型であるとは、共引用テキストによって共引用されている各々の論文が、文中で共通的なものとして扱われている場合のことを指す。例えば、(a)「これまで引用分析研究の分野において、論文間の参照構造を利用し、2論文間の類似度を測るいくつかの手法が提案されてきた^{5),6)}。」^(注3)や(b)「具体的には、引用文章に含まれる語や句を手がかりに引用を分類する試みとして、Garzone²⁰⁾ら、難波ら²¹⁾、Teufel²²⁾らのものがある。」^(注4)のような共引用テキストは共通部分型であると言える。共引用テキストが共通部分型であるということは、共引用している論文の著者から見て、共引用されている論文の組は、ある観点で共通しているということである。文(b)の例であれば、「引用文章に含まれる語や句を手がかりに引用を分類する試み」として、「Garzone²⁰⁾ら」「難波ら²¹⁾」「Teufel²²⁾ら」の3つの論文があることが示されている。これらの論文は、共引用している論文から見て、すべて同じ観点から引用されていると考えられる。

● 差分型

共引用テキストが差分型であるとは、共引用テキストによって共引用されている各々の論文が、文中で異なるカテゴリのものとして扱われている場合のことを指す。例えば、(c)「また、システムと人間との間でインタラクティブにやりとりしながら分類ルールを作成していく手法がPhamら²³⁾によって、有限オートマトンを用いた機械学習による手法がLeら²⁴⁾によって提案されている。」^(注4)や(d)「共引用分析、書誌結合の研究の発展としては、論文の参照関係の分析を論文の検索、分類以外の目的のために利用している研究として論文誌の重要度を測るインパクト・ファクタ関連の研究^{7),8)}、異なる種類の文献の関係をみる研究⁹⁾等がある。」(論文^{7),9)}の組、論文^{8),9)}の組に対する共引用テキストが差分型に相当)^(注5)のような共引用テキストは差分型であると言える。共引用テキストが差分型であるということは、共引用している論文の著者から見て、共引用されている論文の組は、ある観点では共通しているが、目的や手法など、異なる点も含んでいるということである。文(d)

(注3)：論文間の参照情報を考慮した関連論文の組織化[15]より抜粋

(注4)：論文の構成単位に基づいた共引用関係の尺度[13]より抜粋

(注5)：公開されている論文DBからのマクロ情報抽出に対するリサーチマイニング手法と他手法の比較[19]より抜粋

表 1 予備調査に使用した論文と共引用テキストの分類

論文 ID	論文タイトル	発行年	引用文献数	共引用文数	共引用テキスト(個)			計
					共通型 共通部分型	差分型	主従型	
#1	From "Dango" to "Japanese Cakes": Query Reformulation Models and Patterns	2009	27	12	10	5	1	16
#2	Visual Query Suggestion	2009	26	9	9	0	0	9
#3	Structured Query Suggestion for Specialization and Parallel Movement: Effect on Search Behaviors	2012	30	13	12	3	0	15
#4	Context-aware Citation Recommendation	2010	27	6	4	1	3	8
#5	Detecting Topic Evolution in Scientific Literature: How Can Citations Help?	2009	33	11	11	5	2	18
#6	Discovering Aspects of Web Pages from Their Referential Contexts in the Web	2004	16	3	2	1	1	4
#7	Generating Extractive Summaries of Scientific Paradigms	2013	93	16	10	2	6	18
#8	A Visual Mining System for Theme Development Evolution Analysis of Scientific Literature	2010	24	6	4	2	0	6
#9	AspecTiles: Tile-based Visualization of Diversified Web Search Results	2012	23	9	5	5	2	12
#10	Organizing Suggestions in Autocompletion Interfaces	2009	10	2	1	1	0	2
#11	引用箇所の間隔に基づいた共引用の検討	2007	24	6	1	5	0	6
#12	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	1999	20	1	1	0	0	1
#13	公開されている論文 DB からのマクロ情報抽出に対するリサーチマイニング手法と他手法の比較	2004	11	3	1	3	0	4
#14	Web 検索時の行動情報を用いたクエリ修正タイプの予測	2012	26	5	4	1	1	6
#15	論文間の参照関係を考慮した関連論文の組織化	2001	24	7	3	2	0	5
#16	サイエンスマップによる科学研究の動的変化の観測	2009	13	2	2	0	0	2
#17	コミュニティマイニングにおける Web 引用解析と文献引用解析の比較	2004	16	5	3	2	1	6
#18	論文のラベル付きクラスタリングのための情報利得を用いたキーワード選定	2012	9	4	3	1	0	4
#19	論文データベースからの研究トピック抽出	2005	10	4	2	2	0	4
#20	Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見	2006	14	2	1	1	0	2

の例であれば、「論文の参照関係の分析を論文の検索、分類以外の目的のために利用している研究」として、「論文誌の重要度を測るインパクト・ファクタ関連の研究^{7),8)}」の論文「異なる種類の文献の関係をみる研究⁹⁾」の論文の 2 種類があることが示されている。この 2 種類の論文は、共引用している論文から見て、ある同じ観点から引用されているが、手法的には異なるカテゴリのものとして扱われていると考えられる。

(2) 主従型

共引用テキストが主従型であるとは、共引用テキストによって共引用されている各々の論文が、文中で主従の関係にある場合のことを指す。例えば、(e)「実際に既存研究¹¹⁾において、商品購入に関する検索に代表されるトランザクショナル検索²⁾では、他のタスクに比べて広告のクリックスルーレートが上昇するという実験結果が報告されている。」^(注6)のような共引用テキストは主従型であるといえる。この時、共通型とは違い、共引用されている 2 つの論文は並列の関係とならず、一方がもう一方に従う関係になる。文(e)の例であれば、「既存研究¹¹⁾」について説明する際に、「トランザクショナル検索²⁾」を用いているので、後者の論文は前者の論文に従う関係にあると考えられる。他にも、「²⁾は、 $\times \times$ ⁵⁾を用いた手法を提案している。」や、「⁵⁾は、 $\times \times$ ⁷⁾の分野でよく用いられる手法である。」などといった共引用テキストは主従型である。

ここで、共引用テキストの分類は、1 つの文と、それが含む 1 組の共引用に対して定義されることに注意が必要である(す

なわち、同一の文であっても、対象となる共引用の組が異なれば、異なるタイプの共引用テキストとなることがあり得る)。

3.3 共引用テキストの分類による二論文間の関係抽出

我々は、3.2 節で述べた 2 種類の関係の分類について、タイプによって共引用テキストにどのような特徴があるかを調べ、2.2 節で述べた共引用における引用部間の間隔が、二論文間の関係の種類に密接に関わってくるのではないかと考えた。

共引用テキストが同一箇所における列挙共引用を含んでいる場合は、共引用されている各々の論文は共通の観点から引用されているので、この共引用テキストは共通部分型であると考えられる。それに対し、共引用テキストが同一文における共引用を含んでいる場合は、その文における引用の仕方によって、共引用テキストのタイプが異なることがある。例えば、3.2 節における文(b)と文(c)は、どちらも同一文における共引用を含んでいるが、共引用テキストのタイプは異なる。このような場合についても、引用部間のテキスト長や、形態素解析の結果などを用いて、共引用テキストのタイプを判定できると考えているが、具体的な手法の考案については、今後のより深い考察が必要であると考えられる。

4. 予備調査

本章では、3.2 節で述べた関係の分類について、予備調査を行った結果とその考察について述べる。

4.1 調査の概要

共引用テキストから関係を抽出する際に、それぞれのタイプの共引用テキストにどのような文法的特徴がみられるかや、そ

(注6): Web 検索時の行動情報を用いたクエリ修正タイプの予測(表 1 の論文#14)より抜粋

こからどのような関係が抽出できるかという観点を考慮することは重要である。そこで我々は、実際の論文中にどのような共引用テキストが存在し、それらが関係の分類のうちどれに属するかを調べることにした。

調査をする論文は、そこで使われている技術や専門用語などについて、我々が比較的容易に理解できるようなものである必要がある。そこで今回は、文献情報マイニングと検索エンジンについての論文 20 本（うち、英語論文 10 本、日本語論文 10 本）を対象として、そのアブストラクト・本文全文から共引用テキストに相当する文をすべて抜き出し、関連の分類のうちどれに属するかを手動で判定した。

4.2 調査の結果と考察

今回、調査の対象とした論文のタイトル、発行年、引用文献数、論文中における共引用を含む文の数（「共引用文数」と表す）、それぞれの論文の共通型（共通部分型、差分型）、主従型の共引用テキストの数を表 1 に示す。ただし、共引用テキストについて、同一箇所でも 3 つ以上の論文の列挙共引用が存在するなど、数え上げの意義が薄いものについては、すべてまとめて 1 つと数えた。

まず、共引用テキストの数について見ると、引用文献数に比して、概ねどの論文も多くの共引用テキストを持っているといえる。加えて、共引用テキストの分類の分布について見ると、すべての論文において、共通型、特に共通部分型の共引用テキストが比較的多くなる傾向があることがわかる。これは、今回調査した論文の中に、列挙共引用を含む記述が多数見られたことが原因であると考えられる。特に、論文におけるイントロダクションの章や、自らの研究に対する関連研究を紹介する章では、この傾向が強く見られた。これらの章では、自らの研究の背景や、関連する研究を構造化して記載することが多く、あるトピックの研究論文についてまとめて紹介したり、それらの違いを説明したりすることが多くなるため、共通型の共引用テキストが多くなると考えられる。主従型の共引用テキストについては、論文 #7 を除き、あまり見られなかった。主従型には、「 \times^2 」は、 $\times \times^5$ を用いた手法を提案している。」のように、共通型と比べると、若干文構造が複雑になるという特徴があり、論文の著者はそのような執筆の仕方を選んでいるからである可能性がある。また、英語論文と日本語論文での共引用テキストの分布の違いを調べたが、特に違いは見られなかった。

5. ま と め

本稿では、複数の論文を共引用している論文の共引用テキストを用いて、二論文間の関係を抽出する手法について考察した。研究者が論文をサーベイする際に、二論文間の関係を可視化して提示することは、それらの論文の違いや研究分野における位置づけを理解するために重要であるが、直接の参照関係にない二論文の関係を、論文の内容まで考慮して抽出するのは難しいと考えられる。そこで、論文の中から、共引用している論文の両方について言及している箇所である共引用テキストを抽出し、その分類を行うことで、二論文間の関係抽出を行うことを考えた。今後の課題として、共引用テキストの文構造などから、こ

れらを機械的に分類する手法を考案し、実際に論文に適用することで、その有用性を示すことが挙げられる。

謝辞 本研究の一部は科研費 22700097 の助成を受けたものである。

文 献

- [1] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999.
- [2] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on world wide web*, pages 517–526, 2002.
- [3] C. D. V. Hoang and M. Kan. Towards automated related work summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 427–435, 2010.
- [4] V. Qazvinian, D. R. Radev, S. M. Mohammad, B. J. Dorr, D. M. Zajic, M. Whidby, and T. Moon. Generating extractive summaries of scientific paradigms. *Journal of Artificial Intelligence Research*, 46:165–201, 2013.
- [5] 榊剛史, 松尾豊, 市瀬龍太郎, 武田英明, and 石塚満. 論文データベースからの研究トピック抽出. 人工知能学会全国大会, 2005.
- [6] X. Tang, C. C. Yang, and M. Song. Understanding the evolution of multiple scientific research domains using a content and network approach. *Journal of the American Society for Information Science and Technology*, 64(5):1065–1075, 2013.
- [7] J. Wang, C. Wen, S. Wu, and H. Q. Vu. A visual mining system for theme development evolution analysis of scientific literature. *Journal of Digital Content Technology and its Applications*, 4(3):215–223, 2010.
- [8] 難波英嗣 and 谷口裕子. 学術論文データベースからの研究動向情報の抽出と可視化. 言語処理学会第 12 回年次大会, 2006.
- [9] 近藤友樹, 難波英嗣, 奥村学, 新森昭宏, 谷川英和, and 鈴木泰山. 論文データベースからの研究動向情報の抽出. 言語処理学会第 13 回年次大会, 2007.
- [10] M. M. Kessler. Bibliographic coupling between scientific papers. *American documentation*, 14(1):10–25, 1963.
- [11] H. Small. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, 24(4):265–269, 1973.
- [12] 江藤正己. 引用箇所の間隔に基づいた共引用の検討. 第 18 回データ工学ワークショップ (DEWS), pages L1–1, 2007.
- [13] 江藤正己. 論文の構成単位に基づいた共引用関係の尺度. 情報処理学会論文誌, 49(7):1–15, 2008.
- [14] 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, 6(5):43–62, 1999.
- [15] 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, 42(11):2640–2649, 2001.
- [16] S. Teufel, A. Siddharthan, and D. Tiddhar. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 103–110, 2006.
- [17] H. Nanba, T. Abekawa, M. Okumura, and S. Saito. Bilingual presri-integration of multiple research paper databases. In *Proceedings of Recherche d'Information Assistée par Ordinateur*, pages 195–211, 2004.
- [18] Q. He, B. Chen, J. Pei, B. Qiu, P. Mitra, and L. Giles. De-

tecting topic evolution in scientific literature: how can citations help? In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 957–966, 2009.

- [19] 吉田誠, 小林隆志, 横田治夫. 公開されている論文 DB からのマクロ情報抽出に対するリサーチマイニング手法と他手法の比較. *情報処理学会論文誌*., 45(7):24–32, 2004.