

## 論文アブストラクト中の重要文を識別するための手 がかり表現の抽出

大谷, 周平  
九州大学大学院統合新領域学府ライブラリーサイエンス専攻 : 修士課程

<https://hdl.handle.net/2324/1854784>

---

出版情報 : 九州大学, 2014, 修士, 修士  
バージョン :  
権利関係 :

修 士 論 文

論文アブストラクト中の重要文を識別するための手がかり表現の抽出  
Extracting Cue Expressions to Detect the Important Sentences from  
Article Abstracts

統合新領域学府ライブラリーサイエンス専攻修士課程

平成 24 年 4 月入学

(氏 名) 大谷 周平

(ローマ字名) OTANI Shuhei

# 目次

目次 .....	2
<b>1. はじめに.....</b>	<b>4</b>
1.1 研究の概要 .....	4
1.2 研究の背景 .....	4
1.2.1 日常的な情報検索と学術情報検索の違い .....	4
1.2.2 網羅的な学術情報検索の課題 .....	5
1.3 本研究で行う支援 .....	6
<b>2. 関連研究.....</b>	<b>8</b>
<b>3. 提案手法.....</b>	<b>10</b>
3.1 提案手法の考え方 .....	10
3.2 提案手法の手順 .....	11
3.3 提案手法のメリット .....	12
<b>4. 実験.....</b>	<b>13</b>
4.1 実験環境・実験データ .....	13
4.1.1 実験環境 .....	13
4.1.2 手がかり表現抽出用のデータ .....	13
4.1.3 評価用データ .....	14
4.2 実験：手がかり表現の抽出と評価 .....	16
4.3 結果 .....	18
4.3.1 実験結果詳細（10万件） .....	18
4.3.2 抽出用データの増加に伴う評価の推移.....	19
4.3.3 抽出された手がかり表現 .....	20
<b>5. 考察.....</b>	<b>22</b>
5.1 提案手法の有効性 .....	22
5.2 提案手法の課題及び改善 .....	23
5.2.1 誤って重要文と判定されている文.....	23
5.2.2 提案手法により抽出できていない重要文.....	24

6. おわりに.....	26
謝辞 .....	27
付録 .....	28
参考文献 .....	31

# 1. はじめに

## 1.1 研究の概要

本研究では学術情報検索の支援を目的として論文アブストラクト（以下、アブストラクトと表記する）中の重要文を識別する手がかり表現の抽出を行った。アブストラクトのどの文を重要とみなすかという点は、検索者の情報要求によって異なると考えられる。しかし、背景に該当するような文を重要文とみなすことはほとんど考えられず、多くは論文の特徴や学問領域における貢献を表した文、論文の目的・手法・重要な結果のいずれかに該当すると考えた。本研究では、そのなかでも目的文が最も論文の特徴を表していると考え、目的文を重要文とみなしている。重要文には、多くの場合、その文が重要文であることを示すなんらかの手がかり表現、例えば「our objective is」や「in this paper, we show that」のような表現が含まれると考えた。言い換えれば、手がかり表現が含まれている文は重要文である可能性が高いといえる。本研究では手がかり表現は長さ  $N$  の単語列(単語  $N$ -gram)で構成されるものとする。例えば  $N=4$  の時、「our objective is to」や「we aimed to investigate」などを手がかり表現とする。

アブストラクト中の重要文を識別する手がかり表現を網羅的に収集することができれば、手がかり表現を含む重要文も網羅的に抽出することが可能となる。よって、手がかり表現を網羅的に収集するための手法について検討する。

## 1.2 研究の背景

研究者は多大な労力をかけて学術情報検索を行っている。それは日常的な情報検索における典型的な情報要求と異なり、学術情報検索では網羅的な情報検索が求められるからである。まず、この違いについて述べ、その解決のため本研究で行う支援について述べる。

### 1.2.1 日常的な情報検索と学術情報検索の違い

日常的な情報検索においては網羅的な情報検索が求められるケースは頻繁にはない。多くの場合は自分の情報要求に合致する情報をいくつか見つけることが出来ればそれで事足りる。例えば「本日の夕食のお店を探したい」という情報要求に対しては、自分の行動範囲や予算から条件に合致したお店を 1 件ないしは数件発見することができれば十分であろう。このとき、条件に合致する全てのお店を検索する必要性はなく、仮に検索漏れが生じたとしてもそれが重大な損失につながる可能性は極めて低い。このような情報要求では、得られる情報の信頼性や検索ツールが直感的に使えるかといった点が重要視される。

日常的な情報検索が学術情報検索と目的が異なることは、Yahoo Japan が公開する検索キーワードの傾向からもうかがうことができる<sup>1</sup>。頻繁に検索されている検索キーワードのほとんどが「You Tube」や「Facebook」などの Web サービスの名称や、結婚や逝去した著名人の名前である。そこから類推できる情報要求はブックマークの代替として検索データ用いていることや時事情報検索であり網羅性は求められていない。

一方で、学術情報検索では日常的な情報検索と異なり網羅的な検索を求められる機会が多い。統計的データや自分が使用する手法について解説した論文など、日常的な情報検索と同様に情報要求に合致する情報を 1 件ないしは数件見つければよいというケースも存在する。しかし、学術研究では自らが取り組む課題がまだ解決されていないこと、つまり新規性があるものであることを確認する作業が不可欠である。そのためには、同様の目的の研究が行われていないか、一定の範囲において網羅的な調査を行う必要がある。またレビュー論文を執筆する際にも、対象の分野の研究動向を俯瞰的に把握する必要があり、網羅的な情報検索が求められる。そのため、学術情報検索における典型的な情報要求では、情報が網羅的であるかといった点や多少複雑であっても網羅的に情報検索するための機能を備えているかという点が重要視されると考えられる。

このような日常的な情報検索と学術情報検索における典型的な情報要求の違いについて、(諏訪, 2013, p.24) はそれぞれを「発見的探索」と「系統的探索」と位置づけている。そして、「系統的探索」では、全件把握、縮約された全体図の把握、全体の中での相対評価、4 不存在の確認という 4 つの目的が存在し、いずれの目的に対しても条件に合う文献すべての書誌情報が探索の目標となることを述べている。

### 1.2.2 網羅的な学術情報検索の課題

前項で述べた学術情報検索において求められる網羅的な情報検索は相当な労力を必要とする。なぜならば、キーワード検索だけでは的確に自分の情報要求に合致した文献を網羅的に抽出することはできないからである。キーワード検索では、検索者が情報要求を的確に言語化できないケースや、言語化した場合も他の研究者が同一のキーワードを用いているとは限らず関連研究を取りこぼす可能性が常に存在するである。例えば、「文書の取り扱う主題を機械学習で推定したい」という情報要求から「トピックモデル」という検索キーワードにたどりつくのは、事前知識なしでは困難である。加えて「トピックモデル」というキーワードを用いず、上位概念である「自然言語処理」や下位概念である「潜在的ディリクレ配分法 (Latent Dirichlet Allocation)」など個別の手

---

<sup>1</sup> Yahoo! 検索データ. <http://searchranking.yahoo.co.jp/>, (accessed 2015-01-02). では週単位や月単位での検索キーワードの傾向が公開されている。

法だけをキーワードとしている論文が存在した場合、「トピックモデル」という検索キーワードを用いたとしても、これらの論文を取りこぼしてしまうことになる。

そのような検索漏れを避けて、網羅的な情報検索を行うためには、ある程度幅広く論文を抽出し、そこから必要な論文を選別していくことが必要となる<sup>2</sup>。このプロセスは以下のように整理できる。

1. 関連する研究の取りこぼしが無いように、十分に大きな概念を表すキーワードで幅広く論文を取り出す。
2. 検索条件を追加することで検索結果を人手で選別でき、かつ検索したい概念を含んでいると思われる規模にまで絞り込み検索を行う。
3. 検索結果から人手で必要な論文を選別する。
  - A) タイトルをもとに選別を行い、不要なものを取り除く。
  - B) アブストラクトを確認し、不要なものを取り除く。
  - C) 論文本文を確認する。
4. 最終的に残った論文から得られた参考文献やキーワードを用いて、検索を繰り返す。

このプロセスにおいて、特に労力が必要となるのは、3番目の必要な論文を人手で選別していく作業である。タイトル、アブストラクト、論文本文と選別を進めていくにしたがって、研究者の読むべきコンテンツの量は増え、労力は増加していく。重要文を抽出することでこの労力を軽減できると考えられる。

### 1.3 本研究で行う支援

論文の重要文を抽出することによって、アブストラクトを読まずとも不要な論文を選別できる可能性がある。アブストラクトは概ね7から9文程度で構成されているが、その中の重要文だけを読むことで不要な論文を選別できれば、論文の選別にかかる労力を大きく軽減することができる。また、重要文抽出は、前述の網羅的な検索におけるプロセスの1と2で示したキーワード検索や検索結果の絞り込みにも活用することが可能である。

キーワード検索では、検索時に検索範囲を重要文に限定することで、より検索者の情報要求にそった検索結果を得られる可能性がある。検索結果の絞り込みでは、トピックモデルによるクラスタリングに重要文を活用できる。文書の主題を推定するトピックモデルを用いて、類似したトピックの論文をクラスタリングすることは、検索結果の適切

---

<sup>2</sup> 医学のように論文へのキーワード付与・シソーラスの整備・構造化されたアブストラクトなど、網羅的に情報検索を行うことができる環境が整備されている分野は存在する。

な絞り込みを行う上で有効である。しかし、トピックモデルの課題として、機械的に抽出されるトピックは必ずしも人間にとって理解しやすいものとは限らないということが挙げられる。トピックモデルの対象を重要文だけに限定することで、アブストラクト全文や論文本文をトピックモデルの対象とした場合よりも、より人間に理解しやすいキーワードがトピックとして抽出されると考えられる。

以上のようにアブストラクトから重要文を抽出することは、学術情報検索の支援につながると考えられる。本研究では、その前段階としてアブストラクトの各文において、その表現が含まれていれば重要文である可能性が高いといえる手がかり表現の抽出を試みる。



## 2. 関連研究

重要文の抽出に関する研究には多くの蓄積が存在する。本章では特に学术论文を対象とした研究に焦点をあてて、その概要と本研究で取り組むべき課題について整理する。

(徳永ほか, 2007) は、日本語の情報学分野の論文を対象として、アブストラクトにおける重要文を抽出している。まず、重要文を動機文・手法文・結果文に分類し、その上で、アブストラクト中の位置情報、タイトルの単語、手作業で抽出した「本研究」などの手がかり表現の3つを特徴素として用いて重要文抽出を行っている。

(中渡瀬, 大山, 2011) は、情報学や経営工学分野の日本語論文アブストラクトを対象としている。論文の主旨は研究の実施内容(提案や調査, 分析)に多く出現すると述べた上で、「本稿では」や「述べる」など人手で抽出した手がかり表現をもとに、主旨文の抽出を行っている。また、手がかり表現が用いられていない主旨文を抽出するために、手がかり表現から抽出された主旨文で用いられている述語も補足的に手がかり表現として用いている。

(福田ほか, 2013) では、工学系の論文及び特許を対象に、アルゴリズムやツールなどの要素技術とその効果を示す表現の抽出を行っている。人手や機械的に収集した手がかり表現、論文内の位置情報などを要素として機械学習を行い、特許やアブストラクトから要素技術やその効果を示す表現を抽出している。この研究で用いられている手がかり表現は要素技術を表す「用いた」やその効果を表す部分に出現する「向上」などであり、本研究における手がかり表現とは異なるものである。機械的な手がかり表現の収集は、単語の上位・下位の概念にある語の抽出や、構文関係、分布類似度を用いて抽出されている。

(Teufel, Moens, 1998) では、数理言語分野のアブストラクトに対して、人手で作成した「we aim at」や「finally, we show that」などの手がかり表現、文の長さ、位置情報などを用いてアブストラクトの各文の役割を7つに分類することを試みている。この研究は (Teufel, Moens, 2002) で、論文フルテキストを対象とした研究に発展している。アノテーションされた学習用データを用いて目的や背景、研究のオリジナリティなど7つのカテゴリーに分類し、それぞれのカテゴリーで重要な文を抽出している。アルゴリズムは論文における位置情報や、見出し、タイトルの単語、単語の TF-IDF、引用情報を含んでいるかなどの16の要素を用いている。

(Ruch ほか, 2007) では、生命科学分野のアブストラクトを対象としている。構造化されたアブストラクトの情報をもつ生命科学系のデータベース Medline のデータを学習用データとして用いて、生命科学分野で構造化されていないアブストラクトを目的、手法、結果、結論の4つに分類している。学習用データの他にアブストラクト内におけ

る文の位置情報を用いることで精度を上げている。この研究では先行研究に基づいて、結論がもっとも検索において有効な部分であり重要文であると仮定している。しかし、ここで参照されている先行研究（Tbahriti ほか, 2006）では、目的を表す文も結論と同様に検索に有効であると述べられているが、Rush らの研究では、その点については触れられていない。

このように、アブストラクトの重要文抽出に関連する従来研究は、手がかり表現を前もって人手で準備する必要がある手法か、分野が限定される手法である。人手で手がかり表現のリストを作成する場合、さまざまな表現のバリエーションも含めて網羅的に収集することは大変な労力が必要となる。よって、本研究では対象分野を限定しない条件で、機械的に手がかり表現を抽出する手法を提案する。

### 3. 提案手法

本章では、関連研究をふまえて自然科学系分野全般に適用可能で、かつ事前の学習用データ等を必要としない重要文を識別する手がかり表現の抽出方法を提案する。対象とする重要文については、目的を表す重要文がもっとも論文の特徴を端的に表していることが多いと判断したため、目的文を重要文とみなしている。

#### 3.1 提案手法の考え方

多くの場合、論文のタイトルにはその論文がどのような貢献をしたのか、その論文のオリジナリティを表す内容が端的に表現されている。学術論文の書き方の指南書でも、論文のオリジナリティを端的に表したタイトルがよいタイトルとして示されている。例えば（木下, 2002, p.201）では、「表題には「何を」研究したかを具体的に示す。また、「どんな方法で」研究したかも示してほしい。」と述べられている。（石井, 2014, p.22）では、よいタイトルの条件として、論文のメインアイデアが簡潔に表現されていること、タイトルのみで完全に論文のテーマと内容が分かること、研究の範囲を適切に示していること、の3つを挙げている。

タイトルが端的にその論文の特徴を表しているのであれば、アブストラクト中の重要文にもタイトルと共通する表現が多く出現すると考えられる。この性質を用いて、アブストラクト中の文からタイトルと単語の共通する割合が高い文を抽出する。抽出した文の集合には、重要文が多く含まれていると考えられる。ここで抽出したタイトルと単語の共通する割合が高い文を疑似重要文と定義する。疑似重要文の集合には、重要文ではないものも含まれる。また、重要文であっても疑似重要文には含まれないものも含まれる。しかし、疑似重要文の集合に含まれる重要文の割合は、アブストラクトに含まれる重要文の割合よりも高いと考えられる。この全論文アブストラクト中の文の集合・重要

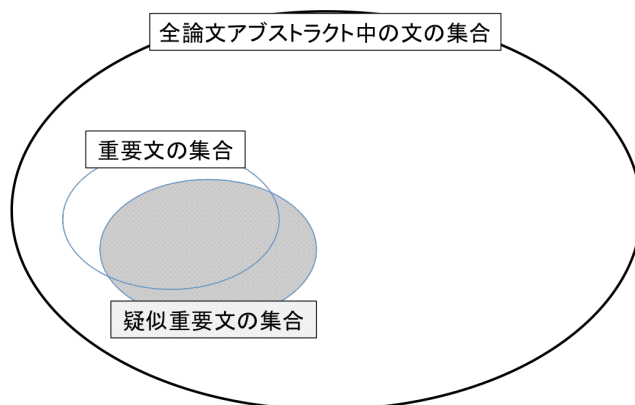


図 1 全論文アブストラクト・重要文・疑似重要文の関係

文の集合・疑似重要文の集合の関係は、図1のように考えられる。この時、重要文を示唆する表現  $ws$  は、重要文に特徴的に出現する表現つまり、

$$\frac{ws \text{ を含む重要文の数}}{ws \text{ を含む文の数}}$$

という割合が高い表現  $ws$  である。もし、重要文と疑似重要文で重なり合う部分が十分に大きければ、この割合を、疑似重要文を利用して以下で近似することができる。

$$\frac{ws \text{ を含む疑似重要文の数}}{ws \text{ を含む文の数}}$$

### 3.2 提案手法の手順

前節の考え方に基づいた提案手法は以下のとおりである、

1. 疑似重要文を抽出する。タイトルの単語集合を  $W_T(A)$ 、アブストラクト内の任意の文  $S$  の単語集合を  $W_S(A)$  とおいたときに、条件式(1)で閾値  $\alpha$  を超える文  $S$  を疑似重要文とみなす。

$$\frac{|W_T(A) \cap W_S(A)|}{|W_T(A) \cup W_S(A)|} \geq \alpha \quad (1)$$

2. 疑似重要文の集合に表れる N-gram の集合を  $C$  とする。 $C$  は手がかり表現の候補の全体集合である。 $C$  中の表現  $ws$  について、 $ws$  を含む文が疑似重要文である割合を求める。ただし、 $ws$  を含む疑似重要文の数を  $n$ 、 $ws$  を含む全アブストラクトの中の文の数を  $m$  として、この割合  $\tilde{r} = n/m$  で推定する。この時、 $m$  が小さい場合に推定誤差が大きくなり、学術論文で重要文に限らず用いられる一般的な表現や疑似重要文に偏って出現するが低頻度で重要文を示唆する手がかり表現とみなすには信頼性に欠ける表現まで手がかり表現として抽出してしまう。これを避けるために以下の区間推定を行う。 $m$ 、 $\tilde{r}$  を用いて表現  $ws$  を含む文が疑似重要文である真の割合  $r$  は信頼区間 95% で以下の式に表される。

$$\tilde{r} - 1.96 \sqrt{\frac{\tilde{r}(1-\tilde{r})}{m}} \leq r \leq \tilde{r} + 1.96 \sqrt{\frac{\tilde{r}(1-\tilde{r})}{m}}$$

よって、表現  $ws$  を疑似重要文に含む数  $n$  と表現  $ws$  を含む全アブストラクトの中の文の数  $m$  の割合  $\hat{r} = n/m$  が、信頼区間 95% で閾値  $\beta$  以上となる下限は以下の条件式(2)で表される。

$$\hat{r} - 1.96 \sqrt{\frac{\hat{r}(1-\hat{r})}{m}} \geq \beta \quad (2)$$

閾値  $\beta$  を満たさない  $ws$  を  $C$  から取り除く。

3. 残った  $C$  を重要文であることを示唆する手がかり表現とみなす.

### 3.3 提案手法のメリット

本手法はタイトルとアブストラクトの単語の共通する割合という簡便な条件で実現可能である. よって, 手作業で手がかり表現を収集する, 事前に学習用データを準備するという作業を行う必要がない. そのため, 分野を問うことなく適用可能な手法である.

タイトルは論文の特徴を端的に表したものであることが多いということを前提としている. そのため, この手法で抽出される情報はタイトルと重複するものであり, 重要文としてタイトルと併せて提示する意義は少ないのではないかという疑問が生じる. しかし, タイトルだけでは十分に内容を推測できない状況として, 以下の 2 つのケースが考えられる.

一つは, タイトルは端的に表す必要があるため, 一部が省略されているケースである. この場合, タイトルと重要文では同一内容を示しているが, 重要文の方がタイトルよりも字数面での制約がないことから, より詳細な情報を得ることができる.

もう一つは, 読者の関心を引くために修辭的なタイトルをつけるケースである. 修辭的なタイトルの例として, (Tomokiyo, Jones, 2001) の「You're not from 'round here, are you?」や (Alexander, 1965) の「A city is not a tree.」などを挙げることができる. 前者は音声データを用いて英語のネイティブと非ネイティブの機械判定を試みる論文であり, 後者は近代の都市建築計画を批判したものである. タイトルだけでこれらの論文の内容を推測することは困難であるが, 例えば前者について, その重要文である「In this paper, we show that naive Bayes classification can be used to identify non-native utterances of English.」という文がタイトルと共に提示されていれば, 論文が目的としている内容を容易に理解することができる. このようにアブストラクトから重要文を抽出することは, 読むべき論文を選択する際の有効な情報源となり得る.

## 4. 実験

本章では提案手法に基づいた実験を説明する。実験では、提案手法に基づいた手がかり表現の抽出を行い、次いで抽出された手がかり表現がどの程度重要文を抽出することができるのか、評価を行った。実験環境について説明を行い、実験に用いたデータ、手がかり表現の抽出実験、およびその評価結果を述べる。

### 4.1 実験環境・実験データ

本実験では実験データとして、手がかり表現抽出用データと抽出した手がかり表現の評価用データの2つを用いて行う。前者は文字通り手がかり表現を抽出するために用いるアブストラクト集合であり、後者は抽出した手がかり表現を用いてどの程度重要文を抽出できるか、抽出した手がかり表現の有効性を評価するために用いる。まず、実験環境について説明したのちに、両データの作成方法について説明する。

#### 4.1.1 実験環境

プログラミング言語 Python version 3.4.1 および Python の自然言語処理ライブラリ Natural Language Toolkit<sup>3</sup>を用いて実験を行った。Python はデータの事前処理や集計など実験全般に、Natural Language Toolkit はアブストラクトの文単位の分割及び N-gram の構築・頻度集計に用いた。

#### 4.1.2 手がかり表現抽出用のデータ

手がかり表現抽出用のデータ作成には、オランダ Elsevier 社の提供する学術情報データベース Scopus<sup>4</sup>を利用した。Scopus は全分野を対象とした世界最大規模の学術情報データベースである。同種のデータベースであるアメリカ Thomson Reuters 社の Web of Science<sup>5</sup>と比較して、アブストラクトの収集が容易なことから Scopus を選択した。

Scopus を用いて、以下の条件を満たすもの論文を手がかり表現抽出データとして抽出した。

1. 文献タイプが「Article (原著論文)」となっていること
2. 刊行年が 2010 年から 2014 年の 5 年間に刊行されていること

---

<sup>3</sup> Natural Language Toolkit — NLTK 3.0 documentation. <http://www.nltk.org/>, (accessed 2015-01-02).

<sup>4</sup> Scopus - <https://www.scopus.com/>, (accessed 2015-01-02).

<sup>5</sup> Web of Science. <http://apps.webofknowledge.com/>, (accessed 2015-01-02).

### 3. 主題分野が自然科学系の論文であること<sup>6</sup>

それぞれ次のような根拠に基づいて決定した。

- 文献タイプ…「Article」に限定したのは、レビュー文献を除くためである。レビュー文献では、タイトルが研究の目的を表すという仮定が成立しないと考えられるため、除外した。
- 刊行年…対象とする論文の刊行年を 5 年間に限定したのは、データ抽出作業を簡便化するためである。単年のデータを用いた場合は、データの収集がさらに容易になるが、データになんらかの偏りが生じる可能性も考慮し、暫定的に 5 年間とした。
- 主題分野…自然科学系の論文に限定したのは、人文系の論文と比較して論文の構造が明確であると仮定したからである。論文の典型的な構造に序論 (Introduction)、手法 (Method)、結果 (Result)、結論 (Discussion) から構成される IMRAD 形式がある。自然科学系では概ねこの構造に則った形で論文が執筆されるが、実験や調査を伴わない研究も多く含まれる人文系では必ずしもこの限りではないと考えられるため対象から除外した。<sup>7</sup>

以上の条件の範囲で、無作為に 3 万件、6 万件、10 万件の 3 つの規模で手がかり表現抽出用のデータを作成した。

#### 4.1.3 評価用データ

評価用データは手がかり表現抽出用のデータと同様に Scopus から収集した。

1. 文献タイプが「Article(原著論文)」となっていること
2. 2010 年から 2014 年の 5 年間に刊行されていること
3. 畜産学分野の論文であること
4. 手がかり表現抽出に利用したデータに含まれていないこと

条件の 1 と 2 については手がかり表現抽出に用いたデータと条件を揃えるために設定した。

---

<sup>6</sup> 論文の主題分野の判定は、Elsevier 社の雑誌主題分類「All Science Journals Classification」に基づいて、その論文が掲載されている雑誌の主題に基づいて決定した。All Science Journals Classification は、Scopus: 収録コンテンツ。

<http://www.elsevier.com/jp/online-tools/scopus/content-overview>, (accessed 2015-01-02). からダウンロードできるタイトル一覧の Excel ファイルに含まれている。

<sup>7</sup>主題分野に関する仮定は、以下の 2 つの研究からも妥当と考えている。(宮田ほか, 2012) や (佐藤ほか, 2013) で、対象とする論文は英語論文と日本語論文という違いはあるが、自然科学系に比べて、人文系では多様な構造の論文が存在することが指摘されている。

条件 3 で畜産分野に限定したのは、以下の理由に基づく。まず、自然科学系の分野である必要があった。加えて重要文を正確に判定するためには、専門的知識を有する人間が判定を行った方が望ましい。そのため、提案手法ではとくに分野を限定する必要はないが、評価用データ作成のコストから本実験では一分野に限定した。畜産分野を選択したのは、自然科学系であること、関連研究で取り扱われている情報工学や生命医学ではないこと、構造化アブストラクト<sup>8</sup>が普及していないことという 3 つの理由に基づく。

条件 4 については、手がかり表現抽出用のデータに評価用データが含まれていた場合、より良い実験結果が得られると考えられる。しかし、現実的には自分の求める論文が手がかり表現抽出用のデータに含まれていることは極めてまれである。そのため、最も厳しい条件として、評価用データが手がかり表現抽出用データに全く含まれていないという設定とした。

以上の条件をもとに抽出した 120 件のアブストラクトを対象に、畜産分野を専攻する大学院博士課程の学生 2 名が、それぞれ 1 文ごとに論文の役割を評価した。文の役割はあらかじめ導入・目的・手法・結果・結論・その他の 6 つを指定した。このうちの目的文を実験における正解文とした。

選択肢となる役割については、アブストラクトのガイドラインを参考に決定した。アブストラクトの構造を標準化したガイドラインとして代表的なものに米国情報標準化機構 (NISO) と国際標準化機構 (ISO) による 2 つが存在する<sup>9</sup>。これらの 2 つのガイドラインにおいてアブストラクトの構成要素は、以下の 5 つと規定されている。目的 (Purpose)、手法 (Methodology)、結果 (Results)、結論 (Conclusions)、その他の情報 (Collateral Information) である。これはアブストラクトの基本的な構成要素として規定されているもので、実際のアブストラクトでは、背景や研究動向などが記述されることがある。そのため、本実験ではガイドラインで規定されている 5 つの役割に背景や研究動向を示す導入を加えた 6 項目でアブストラクトの各文を分類した<sup>10</sup>。

---

<sup>8</sup> 構造化アブストラクトは生命科学系の分野で採用 IMRAD 形式などの構造を明示的に備えたアブストラクトのことであり、畜産分野では採用されていない。

<sup>9</sup> NISO のガイドラインは以下の URL で公開されている。ANSI/NISO Z39.14-1997 (R2009) Guidelines for Abstracts - National Information Standards Organization.  
[http://www.niso.org/apps/group\\_public/project/details.php?project\\_id=32](http://www.niso.org/apps/group_public/project/details.php?project_id=32), (accessed 2015-01-02).

ISO については、以下の URL で概要をみることができる。ISO 214:1976 - Documentation -- Abstracts for publications and documentation.  
[http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=4084](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=4084), (accessed 2015-01-02).

<sup>10</sup> アブストラクトの構造に関する研究として、(Jamar ほか, 2014) がある。その調査でも導入を含めた 6 つの構造を採用している。



評価用データは、全 907 の文で構成されている。文の役割を決定した院生 2 名を以後被験者 A と B とする。被験者 A, B がそれぞれの文に対して付与した役割を集計した。結果は表 1 のとおりである。

被験者 A, B はお互いに相談することなく、独立して文の役割決定作業を行った。被験者 A, B が目的とした文を重要文とみなし、正解文とした。複数人の評価の一致度をはかるコーエンのカッパ係数

(Cohen, 1960) は 0.59 であり、一致度が高いとみなすことができる 0.6 に近い値であるため、妥当と判断した。

また、評価用データを用いて、重要文はタイトルと共通する単語を多く持つという提案手法の前提が妥当であるか検証した。評価用データ全体におけるタイトルとアブストラクト各文の単語が共通する割合の平均は 0.12 である。一方で目的を示す重要文だけに限定した場合は、被験者 A の平均は 0.27, 被験者 B の平均は 0.22 であり、重要文はタイトルと単語が共通する割合が高いという前提は妥当と考えられる。

#### 4.2 実験：手がかり表現の抽出と評価

実験は手がかり表現抽出用のデータ 3 万件, 6 万件, 10 万件を対象に以下の手順を繰り返した。

提案する手がかり表現を抽出する手法の評価は、抽出された手がかり表現一つ一つが妥当な手がかり表現であるか、否かではなく、抽出された手がかり表現の集合を用いて、評価用データからどの程度重要文を抽出することが出来るのかという点で評価する。よって、実験では抽出された個別の手がかり表現についての善し悪しの判断は行わず、ある  $\alpha$ ,  $\beta$  のパラメタ設定で抽出された手がかり表現の集合に対して、評価用データ中の重要文をどの程度の精度と再現率で抽出しているかという観点から評価した。評価方法は以下の手順となる。

1. 抽出された手がかり表現のうち、いずれか一つでも含む文を重要文と判定し抽出する。
2. 手がかり表現によって抽出された文集合を  $X$ , 被験者が重要文と選択した文の集合を  $Y$  とした時に、精度(Precision)と再現率(Recall)は、それぞれ以下の式で表される。

表 1 評価用データの構造

	被験者 A	被験者 B
導入	113	114
目的	51	69
手法	136	122
結果	464	491
結論	121	85
その他	22	27

$$Precision = \frac{|X \cap Y|}{|X|}$$

$$Recall = \frac{|X \cap Y|}{|Y|}$$

実験における評価では、一定の精度を確保した上で再現率を重視した評価を行った。本研究では抽出した重要文に対する最終的な評価は研究者が行うことを想定している。専門分野の研究者であれば提示されている文が重要文であるかどうかは、おそらく容易に判定できるため、重要なことは正確な重要文のみを提示することではなく、重要文と思われるものを網羅的に提示することである。多少、重要文ではない文が重要文として提示されてその文を読む労力よりも、重要文が抽出されておらずアブストラクト全体を読む方がより労力を必要とするからである。よって、精度よりも再現率が重要となる。しかし、再現率だけで評価した場合、ほぼ全ての文を抽出するような条件が採用されては意味がない。よって、本実験では精度 0.8 を条件とし、精度の条件を満たした上で最も高再現率なものを評価の対象とした。また、抽出する手がかり表現の性質として、一つ一つは高精度で重要文を抽出できるが再現率は低いと考えられる。このような手がかり表現を大量に集めることで、再現率を向上させることができると考えているが、現在の実験規模で精度が低かった場合、より大量のデータを用いて手がかり表現の抽出を行った際に、実用に耐えない精度になることが予想される。この観点からも、精度に一定の制限を設けた上で、再現率で評価を行うことは妥当と考えた。

上記の方針に従い、提案手法で抽出された手がかり表現を評価する。提案手法には、 $N$ 、 $\alpha$ 、 $\beta$  という 3 つのパラメタがある。 $N$  は 3 から 5 までを用いている。2-gram は短すぎて信頼性に欠け、6-gram 以上は 3-gram や 4-gram の組み合わせで表現することが出来ると判断したためである。 $\alpha$ 、 $\beta$  に関しては、評価用データの一部を用いて最適な  $\alpha$ 、 $\beta$  を求め、その値を用いて、残りの評価用データで精度・再現率を求める (2-fold cross validation)。ただし、評価用データの規模が小さく、単純に 2 分割しただけでは、分割したそれぞれの評価用データの性質が大きく異なる可能性があり、結果が極端に上下することが予想されるため、分割をランダムに 4 種類用意し、精度・再現率はそのマクロ平均とすることで、評価用データの偏りを吸収した。 $\alpha$  は 0.1 から 0.35 まで、 $\beta$  の値は 0.4 から 0.9 まで、それぞれ 0.05 刻みで変化させた。ただし、先に示した条件式(2)

$$\tilde{r} - 1.96 \sqrt{\frac{\tilde{r}(1-\tilde{r})}{m}} \geq \beta$$

の左辺は母比率の区間推定 (正規分布による近似) の下限である。この近似精度が良いのは  $m$  が 25 以上と言われているため、これを満たさない ws も手がかり表現としては

抽出しない（候補から取り除く）。

手がかり表現抽出用データの規模に応じて精度・再現率がどのようになるか（どの程度の規模であればどの程度の精度・再現率を実現できるか）を調べるため、手がかり表現抽出用データの規模を3万件、6万件、10万件と変えて実験した。以下に、1つの規模の手がかり表現抽出用データに対する具体的な手順を示す。

1. 評価用データを4つのパターンに乱数を用いて並び替えを行い、それぞれ2つに分割した。これを評価用データ1から4と、評価用データ1'から4'とする。
2. 評価用データ1を $\alpha$ 、 $\beta$ のパラメタ決定に用いる。 $\alpha$ 、 $\beta$ の設定ごとに収集された手がかり表現を用いて、手がかり表現を一つでも含む文を重要文と判定する。判定結果と被験者A、Bによる評価を用いて、 $\alpha$ 、 $\beta$ の設定ごとに精度、再現率を求める。評価方針に従い、精度0.8以上で、再現率をもっとも高い $\alpha$ 、 $\beta$ の設定を求める。この $\alpha$ 、 $\beta$ の設定で評価用データ1'に対して、重要文の抽出を行い、精度、再現率を求める。次に、同様の実験を評価用データ1と評価用データ1'を入れ替えた形で実行し、評価用データ1に対して精度、再現率を求める。評価用データ1および1'に対する精度と再現率は文数が異なるため、マイクロ平均を用いる。
3. 評価用データ2から4及び評価用データ2'から4'に対しても実行する。
4. 得られた結果に対して、精度・再現率のマクロ平均を求める。

なお、手がかり表現抽出用のデータの事前処理として、アブストラクトを文単位に分割、アルファベットを小文字へと統一、ハイフン、カギ括弧などの記号類の除去を行っている。冠詞や前置詞などのストップワードについては、予備実験において除去しない方が良好な結果を得られたため、除去していない。また、STEMMINGなど単語の正規化処理も行っていない。

## 4.3 結果

### 4.3.1 実験結果詳細（10万件）

手がかり表現抽出用のデータ10万件における実験結果の一部を表2に示す。評価用データ1と1'、評価用データ2と2'における実験結果である。3万件、6万件を含めた実験結果は付録に示す。 $\alpha$ は0.1から0.2、 $\beta$ は0.4から0.7の範囲の値が用いられている。 $\alpha$ の値が0.25以上、 $\beta$ の範囲が0.75以上の値は本実験ではパラメタとして採用されることは無かった。

表 2 抽出用データ 10 万件の実験結果（一部抜粋）

	パラメタ		抽出された表現数	パラメタ推定用		パラメタ検証用		マイクロ平均	
	$\alpha$	$\beta$		精度	再現率	精度	再現率	精度	再現率
				評価データ 1		評価データ 1'			
被験者A	0.2	0.4	1014	0.86	0.55	0.77	0.40	0.81	0.47
被験者B	0.2	0.4	1014	0.86	0.44	0.85	0.37	0.85	0.40
				評価データ 1'		評価データ 1			
被験者A	0.15	0.7	691	0.80	0.48	0.86	0.55	0.83	0.51
被験者B	0.1	0.7	1055	0.83	0.42	0.86	0.38	0.85	0.40
				評価データ 2		評価データ 2'			
被験者A	0.15	0.6	587	0.82	0.53	1.00	0.40	0.91	0.46
被験者B	0.15	0.6	587	0.82	0.38	1.00	0.36	0.91	0.37
				評価データ 2'		評価データ 2			
被験者A	0.1	0.7	691	0.80	0.48	0.79	0.65	0.79	0.56
被験者B	0.1	0.7	1055	0.83	0.42	0.73	0.46	0.78	0.44

#### 4.3.2 抽出用データの増加に伴う評価の推移

次に、表 3 に手がかり表現抽出用のデータを 3 万件、6 万件、10 万件と変化させた際の結果を示す。被験者 A のデータに対する評価を精度 A、再現率 A とし、被験者 B のデータを精度 B、再現率 B としている。表現数、精度、再現率いずれも 4 つの評価用データの結果からマクロ平均をとったものである。また、それをグラフにしたものを図 2 に示す。単純に手がかり表現抽出用のデータ規模を増やすことで、得られる手がかり表現の数は増加、精度はほぼ 0.85 前後を維持した上で再現率が向上している。6 万件と 10 万件の結果を比較すると、被験者 B に対する結果では、再現率の上昇が若干頭打ちになりつつある傾向が見て取れるが、被験者 A ではまだ上昇傾向にあり、抽出用データの数をさらに増やすことで、ある程度は再現率が向上することが見込まれる。

表 3 抽出用データ単位の評価結果

	表現数	精度 A	再現率 A	精度 B	再現率 B
3 万件	247	0.77	0.38	0.85	0.22
6 万件	534	0.87	0.42	0.85	0.36
10 万件	810	0.84	0.45	0.84	0.37

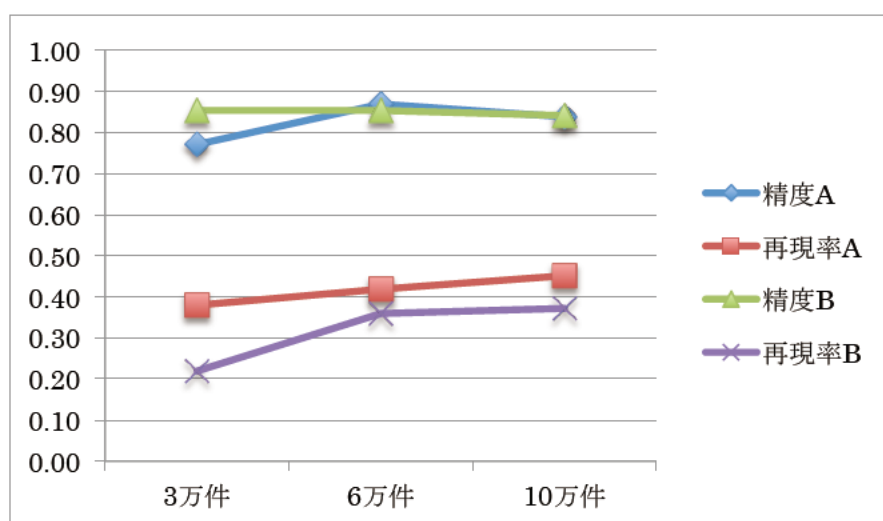


図 2 抽出用データ単位の実験結果

#### 4.3.3 抽出された手がかり表現

最後に抽出された手がかり表現の一部を表 4 に示す。10 万件の抽出データを用い、 $\alpha=0.2$ ,  $\beta=0.4$  の設定の際に抽出されたものである。この時、全体では 1014 の手がかり表現が抽出されている。これは「of this investigation was」と「of this investigation was to」のように N の値によって同一のフレーズから複数手がかり表現が抽出されたものを大量に含んだ数である。各 N の内訳は 3-gram で 431 個、4-gram で 375 個、5-gram で 208 個となっている。

表 4 抽出された手がかり表現（一部抜粋）

3-gram	4-gram	5-gram
a class of	aims to investigate the	developed and validated for the
aim of this	algorithm is proposed for	for the simultaneous determination of
conducted to determine	evaluate the effects of	investigated the effects of
method this paper	model in this paper	is studied in this paper
of this study	study was to assess	objective of the present study

「aim of this」, 「evaluate the effects of」, 「objective of the present study」などは妥当な手がかり表現と考えられる. 4-gram, 5-gram ではこのような表現が大半を占めるが, 「for the simultaneous determination of」のように単独では手がかり表現とはみなしがたいものも含まれる. この傾向は 3-gram でより顕著になる. 「a class of」や「of this study」なども手がかり表現として機能することはあっても信頼性にはかけると考えられる. 実際に評価において「of this study」が精度をさげる原因になっているケースが存在した, 考察でこのような表現の取り扱いについて検討する.

## 5. 考察

本章では、実験結果に基づいて、提案手法の有効性、提案手法の現状の課題さらに改善の方策について考察する。考察では、手がかり表現抽出用のデータとして 10 万件のデータを、パラメタとして  $\alpha=0.2$ ,  $\beta=0.4$  の設定を用いた際に抽出された手がかり表現を用いて検討している。

### 5.1 提案手法の有効性

提案手法の有効性について、2 つの観点から検討する。タイトルとの単語が共通する割合が低い重要文を抽出できているか、人手で手がかり表現を収集する手法に比べて有効かという点である。

まず、タイトルとの単語が共通する割合（以下、単語が共通する割合は  $R$  で示す）が低い重要文を抽出できているかという点である。提案手法は重要文とタイトルで  $R$  の値が高いことを前提としている。この前提については、4 章で妥当であることが確認できた。しかし、提案手法が学術情報検索支援に有効であるためには、 $R$  が低い重要文が存在すること、およびその重要文を本手法で抽出された手がかり表現を用いて判定できていることを確認する必要がある。

手がかり表現を用いて抽出された文における  $R$  の平均は 0.29 とである。これは、評価用データにおける重要文の  $R$  の値と比較しても高い値である（被験者 A で 0.27, 被験者 B で 0.22, 評価用データ全体では、0.12）。 $R$  の値が大きい重要文は手がかり表現を用いた手法でも抽出されやすい傾向にあるといえる。しかし、抽出された個別の文をみると、 $R$  の値が 0.1 前後の重要文の抽出にも複数成功している。これらの重要文は  $R$  の値では抽出できないものであり、手がかり表現を用いることで抽出可能になった重要文といえる。

また、 $R$  を基準に重要文を抽出しようとした場合、例えば  $R=0.3$  で被験者 A のデータに対して重要文を抽出した場合、精度 0.6 で重要文を抽出できるが、一方で再現率は 0.2 にみえない。提案手法により抽出された手がかり表現は、精度 0.85 再現率 0.4 前後と、 $R$  だけを用いた場合に比べて、精度・再現率ともに高い値で重要文を抽出できている点からも提案手法の有効性を確認出来る。さらに  $R$  だけを用いた場合よりも高精度であることから、収集した手がかり表現を用いて新たに疑似重要文を収集することで、 $R$  だけを用いている現状よりも信頼性の高い手がかり表現を得られる可能性も考えられる。

次に人手で手がかり表現を抽出する方法との比較である。人手による手がかり表現の収集は、一定程度収集した段階で新たな手がかり表現を得るために確認しなければなら

ない論文数は飛躍的に増加すると予想される。それに比較して本手法では、ある程度論文数を増やせば再現率が向上することを見込める。また、現在の精度は 0.85 前後であるが、最終的に研究者が目視で確認することを考えれば、精度はもう少し低い値でもそれほど問題にならないと考えられる。精度の目標値を 0.7 程度まで下げることで、それに対応してさらに再現率を向上させることが期待できる。また、評価用データの規模が小さいため、偶然の可能性は否定できないが、提案手法により「was to compare the」という表現が手がかり表現として抽出され、実際に評価用データにおいて重要文の抽出に寄与している。「objective of this paper」などの表現と比較して、「was to compare the」は、人手では手がかり表現としてみなしがたい表現と思われる。このような人手では抽出できない可能性のある手がかり表現を収集できることもある点も提案手法の有効な点であろう。

## 5.2 提案手法の課題及び改善

本節では、現状の提案手法の課題および改善を検討するため、提案手法によって誤って抽出されている重要文と、提案手法によって抽出されていない重要文についてそれぞれ分析を行う。

### 5.2.1 誤って重要文と判定されている文

手がかり表現抽出用のデータとして 10 万件、パラメタとして  $\alpha=0.2$ ,  $\beta=0.4$  を用いた場合、誤って重要文と判定されている文は 3 文である。そこで手がかり表現として用いられているのは「of this study」, 「the proliferation and」, 「the study was」といずれも 3-gram の表現であった。精度をあげることを目的とするならば、単に 3-gram の手がかり表現を除けば良いと考えられる。4 章で示した「a class of」のように手がかり表現として疑わしい表現は 4-gram, 5-gram と比較して、3-gram に多く存在する。しかし、その一方で「of this study」のような手がかり表現でしか抽出できていない重要文も存在しているため、3-gram を除いただけでは再現率が低下してしまい<sup>11</sup>、評価方針で示した再現率を重視する考え方からすると望ましくない。この問題は 3-gram を取り除き、次項で述べる類語辞書や形態素解析を用いて原形を取り出す手法、または抽出用の論文数を増やすことを併用することで解決すると思われる。詳細は後述する。

---

<sup>11</sup> N-gram を 5-gram に限定して、4 章の実験を行った場合、概ね精度は 0.1 向上、再現率は 0.1 低下という結果になり、ほぼ意図したとおりの傾向を示した。



### 5.2.2 提案手法により抽出できていない重要文

提案手法により抽出されていない重要文を分析すると、一切手がかり表現の存在しない重要文というものは、評価用データの範囲では存在しなかった。

現状で抽出されていない重要文の 1/4 は、「objectives of this study」や「this research was aimed」など、本来であれば提案手法で抽出されるべき手がかり表現を含んでいた。これらの表現が抽出されていない原因は類義語や名詞の単数形・複数形、動詞の三人称単数形や過去形などの語形の変化である。今回の実験ではこれらの要素は考慮にいれていないため、「objective」と「objectives」、「study」と「research」、「aim」と「aimed」などを含む表現は、別個に集計されている。そのため、結果としてアブストラクト全集合中の出現頻度が低くなり抽出対象となっていなかった。改善方法としては類語辞書や形態素解析を用いて原形を抽出すること、単純に手がかり表現抽出用のデータ数を増やすことで改善すると考えられる。前項で述べた「of this study」でしか抽出されていない重要文も「objective of this study」は手がかり表現として抽出されているが、「objectives of this study」は抽出されていないことに起因するものであり、この改善案と 3-gram を手がかり表現の候補から外すことで、再現率を維持した状態で精度が向上すると予想される。

提案手法で抽出されていない重要文の残り 3/4 については、何らかの手がかり表現を含んでいるものの現在の手法単独では抽出が困難である。抽出困難な重要文は 2 パターン存在した。

一つは「in order to 動詞」や「to 動詞」と不定詞の目的を表す用法で重要文が示されているケースである。「in order to」という表現は必ずしも論文の重要文のところだけで用いられるというわけではないことや、不定詞には名詞的用法や形容詞的用法もあることから、一律に取り出すことは困難である。英文の構文解析を用いて、不定詞がどのような用法で用いられているか解析することが一つ方法として考えられる。また、目的を表す用法で用いられる際の動詞には「investigate」「understand」「analyze」など一定の傾向が確認される。疑似重要文でよく用いられている動詞を収集し、その動詞を手がかり表現とみなすことで、ある程度このような表現を含んだ重要文も収集することは可能と考えられる。しかし、この場合、1-gram の手がかり表現となるため、精度はかなり低下することが予想される。

もう一つの提案手法では抽出困難な重要文には「here we show that」や「in this paper we provide」などの表現が含まれている。実験前にはこれらの表現は、手がかり表現として想定していたものである。しかし、評価用データを分析するとこれらの表現は論文の結果、結論の部分にも頻繁に用いられており、目的を表す重要文に限定して用

いられている表現ではなかった。実際に抽出用データを用いて、その出現する文の数を調査すると、「here we show that」は、アブストラクト全体では 793 の文に用いられているが、疑似重要文では 145 文にしか用いられていない。

「here we show that」も論文の特徴的な要素を示す手がかり表現であり、検索者にとっては有益であると考えられるため収集の対象とすべきである。提案手法ではこのような表現の抽出は困難であるが、アブストラクトにおける位置情報を用いて、異なる方法で疑似重要文を抽出することにより、抽出できる可能性があると考えている。

## 6. おわりに

本研究では、論文の特徴や学問領域における貢献を表す重要文にはタイトルと共通する単語が多く含まれているという前提のもと、タイトルと単語が共通する割合が高い文を疑似重要文として抽出した。疑似重要文の集合から重要文を示唆する手がかり表現の抽出を行い、抽出された手がかり表現による重要文の抽出性能を評価した。実験により精度 0.85, 再現率 0.4 前後の結果を得られた。この結果は手法の改善や対象とする論文数を増やすことで、一定程度向上することが見込まれ、提案手法の有効性を示すことができた。一方で、提案手法単独では抽出が困難な手がかり表現の存在も明らかになり、網羅的な重要文を抽出するための方策の検討および抽出した重要文を用いた学術情報検索の支援が今後の課題である。

## 謝辞

本稿は筆者が九州大学統合新領域学府ライブラリーサイエンス専攻在籍中に行った研究をまとめたものです。3年間適切に指導いただいた富浦洋一教授に深く感謝いたします。また、さまざまな機会に助言いただいたライブラリーサイエンス専攻の先生がたに感謝いたします。最後に研究にご協力いただくとともに、大学院在籍中にさまざまな支援いただいた九州大学附属図書館，貝塚地区事務部のみなさまにも感謝の意を表します。

## 付録

### 実験結果 (3 万件)

	パラメタ		抽出され た表現数	パラメタ推定用		パラメタ検証用		マイクロ平均	
	$\alpha$	$\beta$		精度	再現率	精度	再現率	精度	再現率
				評価データ1		評価データ1'			
被験者A	0.2	0.4	300	0.85	0.50	0.80	0.28	0.82	0.39
被験者B	0.2	0.45	202	0.83	0.31	0.86	0.16	0.85	0.24
				評価データ1'		評価データ1			
被験者A	0.2	0.4	300	0.80	0.28	0.85	0.50	0.82	0.39
被験者B	0.2	0.4	300	0.80	0.22	0.77	0.31	0.78	0.26
				評価データ2		評価データ2'			
被験者A	0.2	0.5	110	1.00	0.18	0.80	0.16	0.90	0.17
被験者B	0.15	0.7	99	1.00	0.13	1.00	0.14	1.00	0.13
				評価データ2'		評価データ2			
被験者A	0.2	0.45	345	0.82	0.36	0.75	0.35	0.78	0.36
被験者B	0.15	0.65	150	1.00	0.14	0.75	0.13	0.88	0.13
				評価データ3		評価データ3'			
被験者A	0.2	0.4	300	0.83	0.33	0.82	0.43	0.83	0.38
被験者B	0.2	0.4	300	0.83	0.25	0.73	0.28	0.78	0.26
				評価データ3'		評価データ3			
被験者A	0.2	0.4	300	0.82	0.43	0.83	0.33	0.83	0.38
被験者B	0.2	0.45	241	0.80	0.28	0.90	0.23	0.85	0.25
				評価データ4		評価データ4'			
被験者A	0.2	0.45	241	0.80	0.31	0.90	0.36	0.85	0.33
被験者B	0.2	0.5	222	0.80	0.24	0.89	0.23	0.84	0.23
				評価データ4'		評価データ4			
被験者A	0.2	0.4	300	0.82	0.36	0.83	0.38	0.83	0.37
被験者B	0.2	0.45	241	0.90	0.26	0.80	0.24	0.85	0.25

### 抽出された手がかり表現

3-gram	4-gram	5-gram
a study of	evaluate the effect of	of the present study is
a study on	investigate the effect of	study was to investigate the
investigated in this	is proposed based on	study was to investigate the
out to study	main purpose of this	the objective of this work
		the present study
this work reports	this paper presents the	investigated

実験結果 (6 万件)

	パラメタ		抽出された表現数	パラメタ推定用		パラメタ検証用		マイクロ平均	
	$\alpha$	$\beta$		精度	再現率	精度	再現率	精度	再現率
				評価データ1	評価データ1'				
被験者A	0.2	0.45	431	0.92	0.55	0.75	0.31	0.84	0.43
被験者B	0.15	0.6	593	0.87	0.41	0.86	0.32	0.86	0.37
				評価データ1'	評価データ1				
被験者A	0.1	0.8	482	0.91	0.34	0.85	0.50	0.88	0.42
被験者B	0.15	0.7	809	0.80	0.32	0.72	0.41	0.76	0.37
				評価データ2	評価データ2'				
被験者A	0.15	0.65	544	0.90	0.53	0.91	0.40	0.90	0.46
被験者B	0.15	0.6	593	0.82	0.38	0.87	0.36	0.84	0.37
				評価データ2'	評価データ2				
被験者A	0.2	0.45	431	0.83	0.40	0.89	0.47	0.86	0.44
被験者B	0.15	0.6	593	0.87	0.36	0.82	0.38	0.84	0.37
				評価データ3	評価データ3'				
被験者A	0.15	0.65	306	0.89	0.27	1.00	0.38	0.94	0.32
被験者B	0.15	0.6	593	0.83	0.38	0.91	0.34	0.87	0.36
				評価データ3'	評価データ3				
被験者A	0.15	0.6	487	1.00	0.48	0.75	0.40	0.88	0.44
被験者B	0.15	0.6	593	0.91	0.34	0.83	0.38	0.87	0.36
				評価データ4	評価データ4'				
被験者A	0.15	0.6	593	0.87	0.50	0.71	0.40	0.79	0.45
被験者B	0.15	0.6	593	0.93	0.41	0.79	0.31	0.86	0.36
				評価データ4'	評価データ4				
被験者A	0.15	0.65	544	0.82	0.36	0.92	0.42	0.87	0.39
被験者B	0.2	0.45	351	0.85	0.31	1.00	0.35	0.92	0.33

抽出された手がかり表現

3-gram	4-gram	5-gram
analysis in this	aim of this work	for the simultaneous determination
applied to study	evaluate the effects of	intellectual and developmental disabilities
paper deals with	is to investigate the	purpose of this study is
paper studies the	method for the synthesis	the aims of this study
study evaluated the	present study examined the	the present study aimed to
this paper a	the aim of our	this study we investigated the

実験結果 10 万件

	パラメタ		抽出され た表現数	パラメタ推定用		パラメタ検証用		マイクロ平均	
	$\alpha$	$\beta$		精度	再現率	精度	再現率	精度	再現率
				評価データ1		評価データ1'			
被験者A	0.2	0.4	1014	0.86	0.55	0.77	0.40	0.81	0.47
被験者B	0.2	0.4	1014	0.86	0.44	0.85	0.37	0.85	0.40
				評価データ1'		評価データ1			
被験者A	0.1	0.7	691	0.80	0.48	0.86	0.55	0.83	0.51
被験者B	0.1	0.7	1055	0.83	0.42	0.86	0.38	0.85	0.40
				評価データ2		評価データ2'			
被験者A	0.15	0.6	587	0.82	0.53	1.00	0.40	0.91	0.46
被験者B	0.15	0.6	587	0.82	0.38	1.00	0.36	0.91	0.37
				評価データ2'		評価データ2			
被験者A	0.1	0.7	691	0.80	0.48	0.79	0.65	0.79	0.56
被験者B	0.1	0.7	1055	0.83	0.42	0.73	0.46	0.78	0.44
				評価データ3		評価データ3'			
被験者A	0.15	0.65	112	0.89	0.27	1.00	0.38	0.94	0.32
被験者B	0.1	0.7	1055	0.81	0.43	0.82	0.31	0.81	0.37
				評価データ3'		評価データ3			
被験者A	0.15	0.55	1065	0.91	0.48	0.75	0.43	0.83	0.45
被験者B	0.15	0.55	1065	0.82	0.31	0.80	0.40	0.81	0.36
				評価データ4		評価データ4'			
被験者A	0.15	0.55	1065	0.82	0.54	0.71	0.40	0.77	0.47
被験者B	0.1	0.7	691	0.94	0.44	0.75	0.26	0.84	0.35
				評価データ4'		評価データ4			
被験者A	0.2	0.45	562	0.80	0.32	0.83	0.38	0.82	0.35
被験者B	0.2	0.4	645	0.83	0.29	0.92	0.32	0.88	0.30

抽出された手がかり表現

3-gram	4-gram	5-gram
conducted to determine	evaluate the effects of	investigated the effects of
method this paper	model in this paper	is studied in this paper
of this study	study was to assess	objective of the present study
was to estimate	to study the effect	to investigate the influence of
we report the	was to explore the	was carried out to study
with intellectual disability	we report a case	we propose a method for

## 参考文献

1. ANSI/NISO Z39.14-1997 (R2009) Guidelines for Abstracts - National Information Standards Organization.  
[http://www.niso.org/apps/group\\_public/project/details.php?project\\_id=32](http://www.niso.org/apps/group_public/project/details.php?project_id=32),  
(accessed 2015-01-02).
2. Diane Kelly, 上保秀夫, 神門典子, 阿部明典, 加藤恒昭, 清田陽司, 高間康史, 西原陽子, 森辰則. インタラクティブ情報検索システムの評価 : ユーザの視点を取り入れる手法. 丸善出版, 2013. 256p.
3. Imad Tbahriti, Christine Chichester, Frédérique Lisacek, Patrick Ruch. Using argumentation to retrieve articles with similar citations: an inquiry into improving related articles search in the MEDLINE digital library. *Int. J. Med. Inform.*, 2006, vol. 75, no. 6, p. 488–95.
4. ISO 214:1976 - Documentation -- Abstracts for publications and documentation.  
[http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=4084](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=4084), (accessed 2015-01-02).
5. Jacov Cohen. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.*, 1960, vol. 20, no. 1, p. 37–46.
6. Nina Jamar, Alenka Šauperl, David Bawden. The components of abstracts. *New Libr. World*, 2014, vol. 115, no. 1/2, p. 15–33.
7. Patrick Ruch, Celia Boyer, Christine Chichester, Imad Tbahriti, Antoine Geissbühler, Paul Fabry, Julien Gobeill, Violaine Pillet, Dietrich Rebholz-Schuhmann, Christian Lovis, Anne-Lise Veuthey. Using argumentation to extract key sentences from biomedical abstracts. *Int. J. Med. Inform.*, 2007, vol. 76, no. 2–3, p. 195–200.
8. Rashmi Mishra, Jiantao Bian, Marcelo Fiszman, Charlene R. Weir, Siddhartha Jonnalagadda, Javed Mostafa, Guilherme Del Fiol. Text summarization in the biomedical domain: A systematic review of recent research. *J. Biomed. Inform.*, 2014, vol. 52C, p. 457–467.
9. Simone Teufel, Marc Moens. Sentence extraction and rhetorical classification for flexible abstracts, in: *Proceedings of the AAI Spring Symposium on Intelligent Text summarization*, 1998, p. 89–97.
10. Simone Teufel, Marc Moens. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Comput. Linguist.* 2002, vol. 28, no. 4, p.



409-445.

11. Simone Teufel. The structure of scientific articles : applications to citation indexing and summarization. Center for the Study of Language and Information, 2010. 518p.
12. 石井クンツ昌子. 社会科学系のための英語研究論文の書き方 : 執筆から発表・投稿までの基礎知識. ミネルヴァ書房, 2010. 339p.
13. 奥村学, 難波英嗣, 人工知能学会. テキスト自動要約. オーム社, 2005. 159p.
14. 木下是雄. 理科系の作文技術,[改版]. 中央公論新社, 2002. 244p.
15. 佐藤勢紀子, 大島弥生, 二通信子. 学術論文の構造型とその分布 : 人文科学・社会科学・工学 270 論文を対象に. 日本語教育, 2013, no. 154, p. 85-99.
16. 諏訪敏幸. 看護研究者・医療研究者のための系統的文献検索概説. 近畿病院図書館協議会, 2013. 270p.
17. 徳永康次, 延澤志保, 太原育夫. E-032 テキスト構造に着目した学術論文の要旨自動生成のための重要文抽出(E 分野:自然言語・音声・音楽). 情報科学技術フォーラム一般講演論文集, 2007, vol. 6, no. 2, p. 215-216.
18. 中渡瀬秀一, 大山敬三. 論文アブストラクトからの主旨抽出方法. 人工知能学会研究会資料(情報編纂研究会(第 6 回)), No.TETDM-01-SIG-IC-06-03, p. 13-16 (2011)
19. 羽生和紀. 心理学のための英語論文の書き方・考え方. 朝倉書店, 2014. 181p.
20. 福田悟志, 難波英嗣, 竹澤寿幸. 論文と特許からの技術動向情報の抽出と可視化. 情報処理学会論文誌. データベース, 2013, vol. 6, no. 2, p. 16-29.
21. 宮田洋輔, 石田栄美, 池内淳, 安形輝, 上田修一. 学術論文の構成要素と構造. 2012 年度日本図書館情報学会春期研究集会発表要項, 三重大学, 2012-05-12
22. 箕谷千風彦. 推測統計のはなし. 東京図書, 1997. 308p
23. 三輪真木子. 情報行動 : システム志向から利用者志向へ. 勉誠出版, 2012. 205p.