セグメント構造に基づく
学術論文の自動要約

北陸先端科学技術大学院大学
情報科学研究科

辛 沅夏

平成 29 年 3 月

修 士 論 文

セグメント構造に基づく
学術論文の自動要約

1510026　　辛 沅夏

主指導教員　白井 清昭 准教授

審査委員主査　白井 清昭
審査委員　東条 敏
飯田 弘之

北陸先端科学技術大学院大学

情報科学研究科

平成 29 年 2 月

概 要

一般に，研究動向を把握するための先行研究のサーベイは，多くの学術論文を読む必要があり，労力の大きい作業である．そのため，まずは論文の冒頭に書かれた概要を読み，重要な論文を選別してから，より詳しく読むという方法が効率的である．しかし，論文に記載されている概要は簡潔であり，これに書かれている情報だけではサーベイに必要な情報，例えば関連研究との違いや得られた成果などを知ることができないことがある．この場合，論文の本文を読む必要があり，このことがサーベイにかかる負担を大きくしている．

以上のような問題を解決するため，本研究では，サーベイの労力を軽減させることを目指し，学術論文の目的，貢献，関連研究との位置付け，提案手法，評価実験など，論文の主要な要点を全て含む要約を「包括的要約」と定義し，これを自動生成することを提唱する．本論文の提案手法は，重要文抽出型の単一文書要約手法と位置付けられる．従来の単一文書要約手法と異なり，本研究では，学術論文が持つ典型的なセグメント構造に注目し，論文をいくつかのセグメントに分割し，それぞれのセグメントから，各セグメントが持つ特徴を考慮した重要文選択手法で要約を生成し，これらを結合することで最終的な包括的要約を生成する．

本研究で提案する包括的要約の生成手法は以下の通りである．本研究では論文は LaTeX 形式で与えられるものとする．まず，多くの学術論文が共通して持つ典型的なセグメント構造に着目し，論文の章，節または段落を「序論」「関連研究」「提案手法」「実験結果」「結論」の５つのセグメントに分割する手法を提案する．論文の構造を解析し，セグメントに分割する手法として，「節のタイトルを手がかりとする手法」と「関連研究の手がかり句に基づく手法」を提案する．前者の手法では，各セグメント毎に，節のタイトルに出現しやすいキーワードのリストを作成し，このキーワードをタイトルに含む節をセグメントとして抽出する．ただし，「序論」「関連研究」「評価実験」「結論」の４つのセグメントはキーワードのマッチングで抽出し，そのいずれにも該当していない節を「提案手法」のセグメントとして抽出する．次に，「関連研究」のセグメントの抽出率を高めるため，後者の手法を用いる．この手法では，論文を段落に分割した後，「関連研究」のセグメント（段落）の中で典型的に使われると考えられる手がかり句にマッチする段落を抽出することで，「関連研究」に対応するセグメントを抽出する．

次に，抽出したそれぞれのセグメントから重要文を抽出する手法を提案する．学術論文における重要文の現われ方は，論文のセグメントによって異なると考えられる．そのため，それぞれのセグメントに適した重要文抽出手法を開発して重要文抽出に用いる．本論文では，「序論」のセグメントからは論文の目的や貢献について述べている文を，「関連研究」のセグメントからは先行研究と当該論文の研究との差異を説明した文や先行研究の問題点を指摘しつつその論文の提案手法の特色を強調した文を，「提案手法」のセグメントからはその論文の提案手法の概略を説明する文と提案手法の流れや概略などを表す図を，「評価実験」のセグメントからは実験の設定や実験の結果を説明している文や実験の結果

を表す表やグラフを抽出して包括的要約に含める．「序論」のセグメントからの重要文抽出には，セグメント内の文が重要文であるか否かを判定する二値分類器を Support Vector Machine (SVM) で学習し，重要文抽出に使用する．「序論」のセグメントに現れる重要文は論文のアブストラクトにも現れることが多いと考えられる．そのため，「序論」のセグメントに出現する文とアブスラクトに出現する文の類似度を算出し，それが十分に大きいとき，「序論」の文を重要文とみなす．これにより，重要文がタグ付けされた訓練コーパスを自動的に構築し，要約に含めるべきか否かを判定する二値分類器を学習する．機械学習の素性は単語の n-gram(n=1,2,3) を使用する．また，簡単な素性選択を行い，訓練データにおける出現頻度が 1 の素性を削除する．「関連研究」のセグメントからの重要文抽出には，論文の各段落に付与したスコアと TF·IDF スコアを用いてセグメント内の文のスコアを計算し，その上位の文を重要文として抽出する手法を提案する．

本研究で提案する自動要約システムの実装や評価に用いるデータとしては，「言語処理学会論文誌 LaTeX コーパス」における日本語で書かれた論文を使用する．このうち 388 件の論文を学習及び開発データとして使用し，30 件の論文をテストデータとして使用する．実験の結果，節のタイトルを手がかりとした手法では，「提案手法」のセグメントの精度は 83%であったが，それ以外の全てのセグメントの精度は 100%であった．しかし「関連研究」の再現率は 62%と低かった．関連研究の手がかり句に基づく手法では，セグメント抽出の精度は 65%となった．

次に，重要文抽出手法の評価について述べる．「序論」のセグメントから抽出された重要文の精度，再現率，F 値はいずれも 30%程度であった．「関連研究」のセグメントから抽出された重要文の精度，再現率，F 値はそれぞれ 21%，24%，22%であった．また，節のタイトルを手がかりとして検出されたセグメントから重要文を抽出した方が，関連研究の手がかり句によって検出されたセグメントよりも，重要文抽出の再現率が高かった．

「提案手法」「評価実験」のセグメントからの重要文抽出については，本論文では構想を述べただけで，まだ実装が完了していない．各セグメントから抽出した重要文を結合し，元の論文での出現順に並べることで包括的要約を生成する手法もまだ実装されていない．これらの手法を実装することがまず取り組むべき課題である．また本研究で提案した重要文抽出の手法は改善の余地がある．さらに，提案した手法で作成された包括的要約が実際のサーベイにどの程度役に立つのか，すなわち包括的要約の生成が複数の論文の内容を短時間で把握するのにどれだけ貢献するかを確認するための被験者実験も必要である．

目 次

第1章　はじめに

1.1　背景

テキスト自動要約とは，原文書に含まれた情報から重要なものだけを自動的に抽出し，また抽出した情報を簡潔にまとめる処理である．インターネットの普及によって膨大な情報に容易にアクセスできる近年では，入手した情報から必要な情報だけを速やかに選別することが重要である．この際，入手した大量の情報 (テキスト) の全てを読むことは困難である．テキストの要約を自動的に生成し，それをチェックすることで，必要な情報を選別する時間を大幅に短縮できる．このような要求に応じて，自然言語処理の研究分野では自動要約の研究が盛んに行われている．

自動要約では，入力文書が独特な構造や特性を持っている場合には，その特性を利用することで要約の精度を向上できることが知られている [10]．例えば，独特な構造を持つ文書の例として学術論文が挙げられ，学術論文を対象とした自動要約の研究も行われている [1]．

一般に，研究を行う際に先行研究のサーベイは重要であるが，研究のサーベイには，多くの学術論文を読み，その内容を理解する必要があるため，労力の大きい作業である．サーベイは最新の研究動向を把握する際にも行われるが，やはり数多くの論文を読む必要があるため，多大な時間がかかる．このため，サーベイの対象となる全ての論文の全文を読むのではなく，まずは論文の概要だけを読み，概要の内容から重要な論文を選別してから，それらの論文をより詳しく読むという方法が効率的である．このとき，最初に読む論文の概要としては，著者が論文の冒頭に書くアブストラクトや，自動要約の技術を用いて作成した要約が考えられる．

ここで，サーベイに適した要約とは何かを考察する．先に挙げた論文のアブストラクトや自動要約によって生成された要約は，必ずしもサーベイに適していない．論文に記載されているアブストラクトは一般に簡潔であり，これに書かれている情報だけではサーベイに必要な情報，例えば関連する研究との違いや論文の特徴，得られた成果などを知ることができないことがある．このような情報は，「関連研究」や「実験結果」の節に書かれていることが多いが，一般にアブストラクトの文長は制限されているので，アブストラクトには含まれないことも多い．サーベイのために論文を読む場合には，論文の内容をある程度深く理解するため，extended abstract のような長めの要約が必要とされることも多い．

これに対し，自動要約の技術を用いて要約を作成すれば，ユーザが望む長さの要約を生成することができる．ところが，サーベイのために読む要約としては，その論文の目的や

提案する手法の概略だけでなく，先行研究に対する位置付け，実験の設定やその結果，論文の貢献など，論文の主な要点が全て含まれていることが望ましい．従来の自動要約として，元の文書から重要文を選択する重要文抽出型の手法が主流であるが，上記のような論文の要点を全て含んでいるかという観点で重要文が選ばれているわけではない．またサーベイに必要となる内容が含まれているか否かが，重要文を選択する際の基準となってるわけでもない．

このように，先行研究のサーベイの負担を軽減することを目的とした学術論文の自動要約は，そのニーズは大きいものの，これまで十分に検討されていなかった．

1.2 目的

本論文では，多くの学術論文が共通して持つセグメント構造に着目し，学術論文の背景，目的，関連研究との位置付け，提案手法，評価実験など，論文の要点を全て含む要約を「包括的要約」と定義し，この「包括的要約」を自動生成する手法を提案する．セグメント構造とは，ここでは章，節または項によって定義される学術論文の部分テキストの集合と定義する．

本論文の提案手法は，重要文抽出型の単一文書要約手法と位置付けられる．従来の単一文書要約手法 [3, 4, 6, 11] と異なり，本研究は学術論文が持つ典型的なセグメント構造を踏まえて各セグメントが持つ特徴を考慮し，セグメントの種類ごとに異なる重要文選択手法を開発し，それらを適用して重要文を抽出することで包括的要約を作成する．

1.3 本論文の構成

本論文の構成を以下に述べる．2 章では，本論文に関連する先行研究について述べる．要約の手法は，要約対象の数から分類すると「単一文書要約」と「複数文書要約」に分けることができる．また要約文を作成する手法から分類すると，「抽出型要約」と「生成型要約」に分けられる．これらの先行研究を紹介する．3 章では，本論文で提案する包括的要約を生成する手法について述べる．論文のセグメント構造を解析する手法，解析した各セグメントの特徴を考慮した重要文抽出手法，それぞれのセグメントから抽出した要約を統合して包括的要約を生成する手法を説明する．4 章では，提案手法の評価実験について述べる．実験の設定や実験結果について報告する．最後に，5 章では本論文の結論と今後の課題を述べる．

第2章　関連研究

インターネット上には新聞記事，学術論文，e-mail，ブログの日記など様々なジャンルのテキストデータが存在する．テキスト自動要約の分野では，このような異なるドメインの文書から要約を抽出または生成する研究に取り組んでいる．要約を抽出または生成する際には，それぞれのテキストのドメインに合わせた要約手法を導入することで要約の精度を向上させることができると一般的に知られている．入力文書のドメインが特定できれば，そのドメインの文書に共通する固有の特性・文書の構造を利用して要約アルゴリズムを適用することが可能なためである．例えば，新聞記事から要約を作る際には，5W1H(いつ，どこで，だれが，なにを，なぜ，どのように) の内容を漏れ無く含めるべきであり，ブログの日記から要約を作る際には，著者の心境の変化などを解析した上で，そのような情報を要約に含める手法が必要となる．本研究で自動要約の対象とするテキストのドメインは学術論文である．

この章では，テキスト自動要約分野の関連研究について紹介する．要約文を作成する手法は，要約の対象となる文書の数，要約文の生成手法などの基準で分類することができる．要約の対象となる文書の数を基準として分類する場合，一つの文書からその文書の要約を作成する手法を「単一文書要約」という．また複数の文書の内容から要約を作成する手法を「複数文書要約」と呼ぶ．一方，要約の生成手法を基準として分類する場合は，元の文書の中から重要文を選択し，それらをそのまま，あるいは多少の加工を施して抽出し，要約を生成する手法を「抽出型要約」という．また本文の内容を機械に理解させた上で要点をまとめた文を生成させる手法を「生成型要約」と呼ぶ．現時点の自動要約の研究においては，「生成型要約」より「抽出型要約」が主流となっている．しかし，人が文書の要約を作成する際，元の文書から文を抽出するより要点だけをまとめた文を作り出す方が自然である．このように機械が作成する要約も，将来は「抽出型要約」から「生成型要約」に変化していくものと思われる．

以下の節では，抽出型の「単一文書要約」と「複数文書要約」，また「生成型要約」についての研究を紹介する．

2.1　単一文書要約

本研究は重要文抽出による単一文書要約の研究と位置づけられる．以下では，その関連研究について述べる．

Edmundson らは従来よく用いられてきた単語頻度と位置情報という属性に加え，cue words と skeleton という概念を導入することで，重要文抽出型自動要約の典型的手法を確立した [4].

Kupiec らは，Edmundson らの研究を踏まえ，文の長さや大文字の出現を素性として学習した Naive Bayes 分類器によって，原文書におけるそれぞれの文を要約に含めるか否かを分類する手法を提案した [6].

Lin らは，テキストは予測可能な文書構造を持っており，中心的な役割をする文は特定の場所に位置するという考え方に基づき，文のテキスト上における位置によって重みをおく手法 (position method) を研究し，さらにはテキストのドメインによって位置情報の利用を最適化する手法 (optimal positionpolicy) を提案した [8]．しかし，一般にテキストの典型的な文書構造はドメインによって異なるため，position method は一般性を持たないという問題がある．

また，Lin らは，重要文抽出のための学習素性は互いに独立でないという考え方に基づき，先行研究で採用された Naive Bayes でなく決定木を用いて重要文か否かを判定する分類器を学習した [7]．しかし，評価実験の結果，この手法はトピックによっては Naive Bayes よりも優れた成果は挙げられなかった．Lin らはこの理由をテキストのドメインによっては学習素性が実は互いに独立であるためと説明した．

原文書におけるそれぞれの文を要約に含めるかを独立に分類するこれまでの手法とは異なり，Conroy らは，HMM を用いて複数の文の中から要約として抽出するべき文を同時に選択する手法を提案した [3]．彼らの提案手法は，テキストにおける文の位置，文中の単語の頻度，単語のスコアの三つの素性を利用し，さらに文と文の間の局所依存性を考慮して，元の文書から複数の重要文を同時に選択するモデルを提案している．

文章中の文や句の間の役割や関係を表わす談話構造を利用して重要文抽出の精度を高める研究がある．Louis らは，文や句の間の関係から導かれるテキストのグラフ構造とそれらが持つ意味を評価した [9]．彼らはグラフ構造は重要な文を選択する上で重要な指標となり，グラフ構造を重要文抽出に活用する際にその意味属性を有効に使えると主張した．また，Hirao らは談話構造 (rhetorical structure) に基づいた discourse tree から文と文の依存性を表す木構造を生成し，要約生成に利用している [5].

2.2 複数文書要約

ウェブなどに存在する多数のテキストデータから必要な情報を効率的に収集するために用いられるのが「複数文書要約」である．この手法を用いれば，例えば使用者があるトピックをクエリーとして与えたとき，そのトピックに関する最新の新聞記事を集め，要約として表示することなどが可能である．複数文書要約の手法としては，sentence clustering などの手法によって複数の入力文書から類似している部分テキストを見つけ，それから要約を生成する手法が広く使われている [1, 2].

また，学術論文を対象とした複数文書要約手法では，複数の学術論文から，要約対象の論文から他の論文を引用している文の情報や，要約対象となる論文の文を引用している他の論文の情報を利用して要約を生成する研究も存在する [1]．この手法は，論文を引用している文には，その論文の特徴，目的，成果などが書かれている場合が多く，要約に含めるべき文として有用であるという観察に基づく．

2.3 生成型要約

2.1 節，2.2 節で述べた研究は主に「抽出型要約」を生成する手法であった．本節では，「生成型要約」について述べる．

抽出型要約手法で選択した文を最終的な要約にそのまま含めると，場合によっては，前後のコンテキストが考慮されず，その文だけでは理解できない句や代名詞などが含まれることが多い．また，重要文として抽出した文に，重要でない内容も冗長に書かれていることもある．このような問題を，「文圧縮 (compression)」や「文融合 (sentence fusion)」を通して解決しようとする研究がある．「文圧縮」は選択された文から必要のない句や節を除去する作業であり，「文融合」とは入力文書の複数の文から必要な部分だけを切り出し，切り出した部分を貼り付けて要約を生成する手法である [10]．Zajic らは，入力文から導かれる複数の圧縮文の候補から最も適切なものを選択するための multi-candidate reduction フレームワークを提案した [12]．Barzilay は，複数の入力文書中の文から作られる構文木と類似語を用いて，入力文書間で共通する部分テキストを検出し，これらの部分テキストを融合して要約を生成する手法を提案した [2]．

2.4 本研究の特色

本研究の特色は，研究のサーベイに利用することを想定し，論文の主要な要点を全て含む包括的要約を生成する点にある．これまで，学術論文を対象とした自動要約の研究も行われてきたが，サーベイに役に立つという観点で要約の生成を試みた研究はなかった．また，学術論文は一般に典型的な構造を持つ．すなわち，標準的な論文は，序論，関連研究，提案手法，評価実験，結論から構成される．談話構造を考慮した自動要約生成手法 [5] は提案されているが，このような学術論文の構造を考慮して要約を生成する研究はこれまで行われていない．本研究では，学術論文を主要な構成要素（セグメント）に分割した後，それぞれの構成要素から重要文を抽出することで包括的要約を生成する．このとき，構成要素毎に適した重要文抽出手法を適用することで，精度の高い要約を生成する．さらに，要約を構成要素毎に整理し，表形式でわかりやすく提示することで，研究のサーベイに適した要約の生成を目指す．

第3章　提案手法

本章では，学術論文から包括的要約を自動生成する手法について述べる．

3.1　概要

一般に，論文は LaTeX, Word, HTML など様々な形式で作成される．本研究では，論文として LaTeX のソースファイルが与えられるものと仮定する．論文の LaTeX ファイルは常に入手できるわけではないが，現在も多くの論文は LaTeX 形式で作成されている．LaTeX ファイルの特徴としては，章や節などの構造が明示的にマークアップされているため，論文のセグメント構造を解析しやすいという利点がある．

本研究では，LaTeX ファイルの中のコマンドはセグメント構造の解析のみに利用する．具体的には節のマークアップ (\section, \subsection)，参考文献のマークアップ (\cite)，およびアブストラクトのマークアップ (\jabstract [1]) のみを使用する．したがって，LaTeX 以外のフォーマットで書かれた論文においても，節のタイトルや境界，参考文献の参照箇所，アブストラクトの領域が同定できれば，本論文の提案手法を適用することが可能である．また，図や表などの参照コマンド (\ref) や表を作成するコマンド (\begin{tabular}) を検出し，重要な図や表を選択して包括的要約に含めることも可能である．図や表は論文の内容をすばやく理解するのに役に立つと考えられる．

提案手法の処理の流れを図 3.1 に示す．まず，原論文のセグメント構造を解析する．本研究では，学術論文の典型的な構造を考慮し，論文の章または節を「序論」「関連研究」「提案手法」「実験結果」「結論」の 5 つのセグメントに分割する．

[1] このコマンドは，評価実験に用いた「言語処理学会論文誌 LaTeX コーパス」の中で，論文冒頭のアブストラクトをマークアップするために独自に定義されたものである．このコーパスの詳細は 4.1 節で説明する．

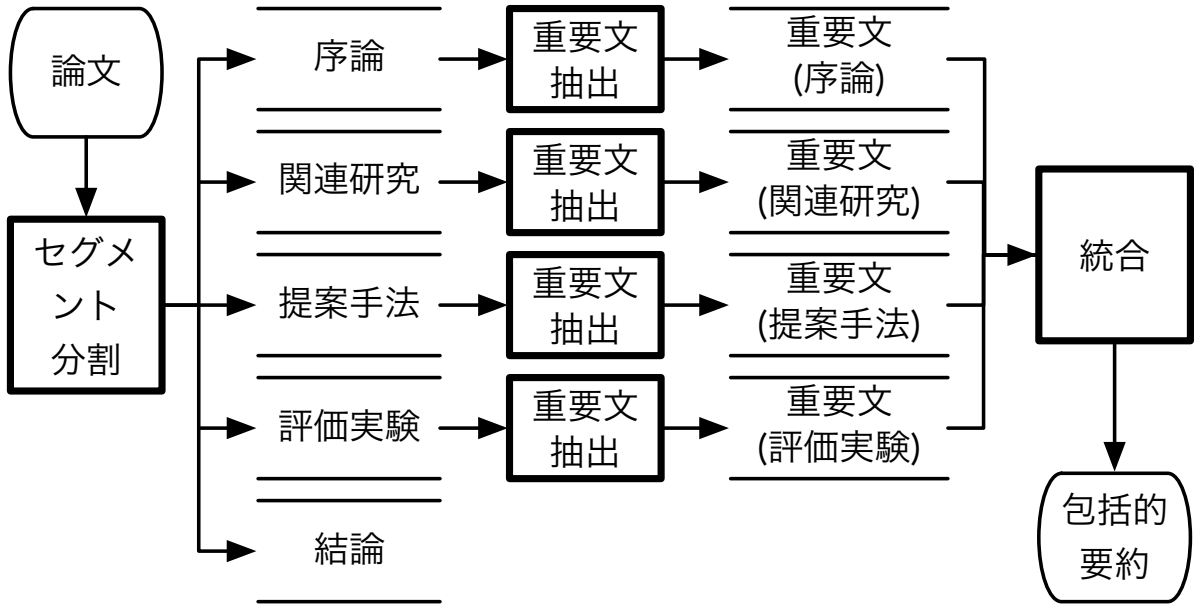

図 3.1: 提案手法の概要

次に，分割したそれぞれのセグメントから重要文を抽出する．重要文の現われ方は，論文のセグメントによって異なると考えられる．そのため，それぞれのセグメントに適した重要文抽出手法を開発し，それを重要文抽出に用いる．ただし，「結論」のセグメントからは重要文抽出を行わない．その理由を以下に述べる．「結論」セグメントに当てはまる章や節では，主に論文のまとめや今後の課題などが書かれていることが多い．しかし，論文のまとめは多くの場合「序論」の内容と重複しており，また今後の課題はサーベイのための要約に含める必要性が低いと本論文では考える．

最終的に，「序論」「関連研究」「提案手法」「評価実験」のそれぞれから抽出された重要文を統合することで，包括的要約を生成する．このように，論文における複数のセグメントからそれぞれ適した手法で重要文を抽出することで，論文の主要な内容を全て含む包括的要約を作成する．

以下，提案手法の詳細について述べる．ただし，「提案手法」と「評価実験」のセグメントからの重要文抽出，ならびに「統合」モジュールは実装が完了していない．そのため現時点での構想を述べる．

3.2 セグメント構造の解析

既に述べたように，本研究では，論文を「序論」「関連研究」「提案手法」「評価実験」「結論」の 5 つのセグメントに分割する．それぞれのセグメントの定義を以下に示す．

- 序論

  論文の最初に書かれ，研究の背景，目的，論文の概要を述べているテキスト

- 関連研究

  先行研究について説明したり，先行研究とその論文の研究との違いを議論しているテキスト

- 提案手法

  その論文が提案する手法の詳細を説明したテキスト

- 評価実験

  論文の提案手法の有効性を評価するための実験について説明したり，その実験結果の考察について述べているテキスト

- 結論

  論文の最後に書かれ，その論文の成果や貢献をまとめたり，今後の課題について議論しているテキスト

セグメント構造の解析は2つの手法によって行う．一つは節のタイトルを手がかりとする手法，もう一つは関連研究の手がかり句に基づく手法である．

3.2.1 節のタイトルを手がかりとする手法

セグメント構造解析の第1ステップとして，節のタイトルに対するパターンマッチングによってセグメントを検出する．

まず，与えられた論文に対し，それを節に分割し，また節のタイトルを検出する．LaTeXのソースファイルでは，論文は \section を境界として節に分割されているとみなせるため，\section によって論文を節に分割する．正確には，\section が分割された節の先頭になるように分割する．さらに，\section{ ... } のように括弧で囲まれた文字列を節のタイトルとして抽出する．なお，LaTeX 以外のフォーマットのファイルを入力とするときも，節の区切りを検出したり，節のタイトルを抽出することは比較的容易に実現できると考えられる．

次に，検出した節が5つのセグメントのどれに該当するかを判定する．「提案手法」を除くセグメントについては，そのセグメントの節のタイトルによく使われると思われる表現をキーワードのリストとしてあらかじめ用意する．そして \section という LaTeX コマンドでマークアップされているテキストにこのキーワードが含まれている場合，その節を該当するセグメントとして分類する．セグメントのキーワードのリストを表 3.1 に示す．これらのキーワードは人手で選定した．

表 3.1: セグメントのキーワードの一覧

セグメント	キーワード
序論	はじめに, まえがき, 序論, はしがき, 背景, 緒論
関連研究	関連研究
評価実験	実験, 評価, 評価実験, 評定実験
結論	考察, 結論, おわりに, 終わりに, 結び, むすび, まとめ, あとがき

「提案手法」のセグメントについては，様々な単語がタイトルに出現する可能性があり，これらの単語や手がかり句をあらかじめ網羅的に収集することが難しいことから，キーワードのマッチングではセグメントの分類をしない．代わりに，他の 4 つのセグメントを同定した後，そのいずれにも該当していない節を「提案手法」のセグメントとみなす．

以上の手法で抽出されるセグメントは節を単位とする．また，論文の中には 6 つ以上の節が存在することがあるので，複数の節が一つのセグメントを構成することがある．また，「提案手法」以外のセグメントについては，表 3.1 に示したキーワードにマッチするタイトルの節が存在しないときは，そのセグメントは検出されない．例えば，「関連研究」というキーワードを含むタイトルを持つ節がひとつも存在しないに論文については，「関連研究」のセグメントは検出されない．

図 3.2 は，本項で提案した手法により論文の節を各セグメントに分類した例である．この図では，対象論文のコーパスにおける番号，タイトル，著者，節のタイトルとセグメント分類結果，そしてその節の下位にある項 (subsection) のタイトルを表示している．segment type が検出されたセグメントの種類を表わす．ここで使われている記号の意味を以下に記す．

`introduction`: 「序論」のセグメント
`related_work`: 「関連研究」のセグメント
`proposed_method`: 「提案手法」のセグメント
`experiment_result`: 「評価実験」のセグメント
`conclusion`: 「結論」のセグメント

全ての節は正しいセグメントに分類されたものの，「関連研究」のセグメントが抽出されていない．

```
文書 ID   : V01N01-03
タイトル : 並列構造の検出に基づく長い日本語文の構造解析
著者      : 黒橋 禎夫, 長尾 眞
------------------------------------------------------------
segment type     : intro
section title    : はじめに
------------------------------------------------------------
segment type     : proposed_method
section title    : 並列構造の検出と文の簡単化
subsection title : 並列構造の検出の概要
subsection title : 並列構造間の関係の整理による文の簡単化
subsection title : 違反関係にある並列構造の修正
------------------------------------------------------------
segment type     : proposed_method
section title    : 係り受け解析
subsection title : 係り受け解析の概要
subsection title : 一定範囲内の文節列の係り受け解析
subsection title : 並列構造の範囲の延長
subsection title : 係り受け解析を失敗した場合
------------------------------------------------------------
segment type     : experiment_result
section title    : 文解析の結果とその評価
subsection title : 定量的評価
subsection title : 関連研究
subsection title : 解析の誤り
------------------------------------------------------------
segment type     : conclusion
section title    : おわりに
------------------------------------------------------------
segment type     : related_study
section title    : related_study
```

図 3.2: 節のタイトルを手がかりとしたセグメント抽出結果

3.2.2 関連研究の手がかり句に基づく手法

4.1 節で後述する 388 件の論文から構成される開発データを用いた予備実験では，「関連研究」のセグメントについては，3.2.1 項で提案した手法では，セグメントが検出されないことが多かった．その原因を分析した結果，比較的最近の論文では，その論文に関連する研究の紹介とその論文の特徴の説明に一つの独立した節を割り当てるケースが多かったが，古い論文ではそのような内容が一つの独立した節として記述されているのではなく，別の節に含まれていたり，節の中の項に割り当てられているケースが多かったことが原因と考えられる．すなわち，節を単位としたセグメントの検出手法では，関連研究に言及したセグメントを見つけることができない論文も多いことがわかった．一方，先行研究とその論文の提案手法の違いを述べている文は，その論文の特徴を理解する上で重要であり，包括的要約に含めるべきである．そのため，「関連研究」のセグメントが検出できないことは重要な問題である．この問題を解決するため，3.2.1 項で述べた節のタイトルに対するパタンマッチで「関連研究」のセグメントを検出できなかったとき，手がかり句を用いて検出する手法を提案する．

関連研究に関する内容がひとつの節として論文中に割り当てられていない場合でも，論文中のどこかには関連研究に言及した一連の文が存在すると仮定する．また，関連研究に言及するのはひとつの文ではなく，ひとつまたは複数の段落によって言及されることが多いと考えられる．そのため，ここでは段落を単位としてセグメントを抽出する．

まず，論文を段落に分割する．LaTeX のソースファイルでは，段落は空行でマークアップされるため，空行を段落の境界とする．また，\section や \subsection のような節や項をマークアップするコマンドも段落の境界とする．

次に，「関連研究」のセグメントに対応する段落を検出する．先ほど述べたように，この手法では関連研究の手がかり句を用いる．手がかり句とは，「関連研究」のセグメントの中で典型的に使われると考えられる表現にマッチするパターンである．手がかり句（パターン）の定義を図 3.3 に示す．

P1: われわれ | 我々 | 本 (研究|手法|論文|稿) | 特徴 | 具体
P2: これ (まで|ら) の (研究|手法|方法) | 提案 | 比較 | 研究 | 方法 | 手法 | CITE
P3: しかし | 一方 | ただ | 違い | 異なる | 異なり | (で|て)(ε|は) ない | いない | できない | でき (る|た)

図 3.3: 「関連研究」のセグメントを検出するための手がかり句

図 3.3 の P1, P2, P3 はそれぞれ，以下の文にマッチすることを想定している．

P1 論文の特徴を述べる文

P2 先行研究との比較を述べている文

P3 先行研究の問題点を指摘する文

また，P2 における CITE は，論文を引用する LaTeX のコマンド \cite にマッチすることを意味する．

図 3.4 は，それぞれ P1, P2, P3 の手がかり句にマッチした文の例を示している．下線はそれぞれのパターンにマッチした箇所を表わす．P2 にマッチした文は，関連研究や関連研究との位置付けについて述べている文ではないため，パタンマッチによる関連研究に関する文の検出に失敗しているといえる．

文書 ID　：V10N01-01
タイトル：日本語固有表現抽出の難易度を示す指標の提案と評価
著者　　：野畑 周，関根 聡，辻井 潤一

（P1 にマッチした文の例）
本研究 では，固有表現抽出の難易度を，テストコーパス内に現れる固有表現，またはその周囲の表現に基づいて推定する指標を提案する．

（P2 にマッチした文の例）
Bagga et. al [CITE] は，MUC で用いられたテストコーパスから意味ネットワークを作成し，それを用いて MUC に参加した情報抽出システムの性能を評価している．

（P3 にマッチした文の例）
あらゆるコーパスを統一的に評価できるような，固有表現抽出の真の難易度は，現在存在しないので，今回提案した難易度の指標がどれほど真の難易度に近いのかを評価することは できない．

図 3.4: 「関連研究」の手がかり句にマッチした文の例

ある段落の中に，図 3.3 のパターンにマッチする文があれば，その段落，及びその前に出現する 2 つの段落を「関連研究」のセグメントとして抽出する．手がかり句にマッチした段落だけでなく，前の 2 つの段落まで抽出した理由は以下の通りである．開発データの論文を調査したところ，関連研究に関するセグメントは，複数の段落で構成されることが多かった．また，多くの論文では，3 つの段落で関連研究について述べていた．また，P1, P2, P3 のパターンは，関連する研究について紹介する段落を検出するために定義したのでなく，関連研究についての紹介が一通り終わった後，関連する研究から位置づけられる当該研究の特徴や意義などが書かれている段落を検出するために用いている．すなわち，P1, P2, P3 にマッチする段落は関連研究を紹介する一連の段落の最後であることを想定している．

複数の段落が図 3.3 のパターンによって「関連研究」セグメントとして検出された場合は，その中からもっとも適切な段落を選択する．段落のスコアは式 (3.1) のように定義

する．

$$Score = n_{P1} \cdot s(P1) + n_{P2} \cdot s(P2) + n_{P3} \cdot s(P3) \tag{3.1}$$

$$s(P1) = 10, \quad s(P2) = 3, \quad s(P3) = 2 \tag{3.2}$$

n_{P1}, n_{P2}, n_{P3} は，それぞれパターン P1, P2, P3 がマッチした回数である．一般に，一つの文に対し，P1, P2, P3 のパターンが複数回マッチすることがあるが，そのときはマッチした回数だけ n_{P1}, n_{P2}, n_{P3} を加算する．一方，$s(P1)$, $s(P2)$, $s(P3)$ はそれぞれ P1, P2, P3 に対して与えられるパターンのスコアである．これは式 (3.2) のように定義する．すなわち，P1, P2, P3 の順に高い重みを与える．また，10, 3, 2 というスコアの大きさは，開発データの論文を参照し，これらから関連研究の段落が適切に抽出されるように定めた．以上をまとめると，段落中に含まれる文が P1, P2, P3 のパターンにマッチする度に 10, 3, 2 という点を与え，その点を合算したものを段落のスコアとする．

図 3.5 は，本項で提案した手法により，「関連研究」のセグメント（段落）を抽出した例を示している．3 番目の段落のスコアは，論文中の全ての段落の中で最も高かった．そのため，この段落と，この段落の上の 2 つの段落が「関連研究」のセグメントとして抽出されている [2]．なお，3 番目の段落では，手がかり句がマッチした箇所を下線で表している．

[2]提案システムでは，\subsection でマークアップされる節のタイトルも 1 つの段落として取り扱う．この例では，「\subsection{ 固有表現抽出の難易度における前提 }」が 1 つの段落となっている．

文書 ID　：V10N01-01
タイトル：日本語固有表現抽出の難易度を示す指標の提案と評価
著者　　：野畑 周，関根 聡，辻井 潤一

\subsection{固有表現抽出の難易度における前提}

異なる分野における情報抽出タスクの難易度を比較することは，複数の分野に適用可能な情報抽出システムを作成するためにも有用であり，実際複数のコーパスに対して情報抽出タスクの難易度を推定する研究が行われてきている．
Bagga et. al [CITE] は，MUC で用いられたテストコーパスから意味ネットワークを作成し，それを用いて MUC に参加した情報抽出システムの性能を評価している．
固有表現抽出タスクに関しては，Palmer et. al [CITE] が Multilingual Entity Task [CITE] で用いられた 6 カ国語のテストコーパスから，各言語における固有表現抽出技術の性能の下限を推定している．

本研究 では，固有表現抽出の難易度を，テストコーパス内に現れる固有表現，またはその周囲の表現に基づいて推定する指標を 提案 する．
指標の定義は，「表現の多様性が抽出を難しくする」という考えに基づいている．
文章中の固有表現を正しく認識するために必要な知識の量に着目すると，あるクラスに含まれる固有表現の種類が多ければ多いほど，また固有表現の前後の表現の多様性が大きいほど，固有表現を認識するために要求される知識の量は大きくなると考えられる．
あらゆるコーパスを統一的に評価できるような，固有表現抽出の真の難易度は，現在存在しないので，今回 提案 した難易度の指標がどれほど真の難易度に近いのかを評価することは できない．
本論文 では，先に述べた，「複数のシステムが同一のコーパスについて固有表現抽出を行った結果の評価」を真の難易度の近似と見なし，これと 提案 した指標とを比較することによって，指標の評価を行うことにする．
具体的 には，1999 年に開かれた IREX ワークショップ [CITE] で行われた固有表現抽出課題のテストコーパスについて 提案 した指標の値を求め，それらと IREX ワークショップに参加した全システムの結果の平均値との相関を調べ，指標の結果の有効性を検証する．

図 3.5: 「関連研究」の手がかり句にマッチした段落の例

3.3 重要文抽出

セグメント構造の解析後，包括的要約を生成する次のステップとして，それぞれのセグメントから重要文を抽出する．ここでの重要文とは，包括的要約に含めるべき文であり，サーベイの際に論文の内容を把握することのできる文と定義する．

重要文抽出による要約生成の一般的な方法は，テキストに含まれる文に対し，その文の重要度を求め，重要度の高いいくつかの文を選択するというものである．文の重要度は，その文に含まれる単語の頻度，同じく文に含まれる単語の TF·IDF スコア，文の位置情報，重要文に出現しやすい手がかり句の有無，などを基に計算されることが多い．また，自動要約では要約率の制御が重要である．要約率とは，自動的に作成された要約の長さと元のテキストの長さの比である．要約率はユーザが指定するので，自動要約システムは様々な要約率に対して要約を生成することが求められる．重要文抽出による自動要約では，選択する文の数を変更することで要約率を柔軟に変更できる．すなわち，重要度の高い順に文を選択していき，要約がユーザが指定した要約率を満たした時点で文の選択を止めればよい．

本研究では，「序論」「関連研究」「提案手法」「評価実験」のセグメントから重要文を抽出する．その際，セグメントによって重要文の現われ方や内容が異なると考えられる．例えば，重要文に出現する手がかり句の有無によって重要文の重要度を計算する際，その手がかり句はセグメントによって異なるだろう．「序論」と「関連研究」のセグメントの重要文に出現しやすい表現の例を表 3.2 に示す[3]．以上の考察を踏まえ，本研究では，セグメントに出現する重要文の性質を考慮して，セグメント毎に適した重要文抽出手法を提案する．以降の項では，「序論」「関連研究」「提案手法」「評価実験」のそれぞれから重要文を抽出する手法を順に説明する．

表 3.2: 特定のセグメントの重要文に出現しやすい表現の例

セグメント	出現しやすい表現
序論	本論文の目的は～
関連研究	先行研究との違いは～

3.3.1 「序論」からの重要文抽出

「序論」のセグメントでは，論文の目的を説明している文，論文の貢献を説明している文などがサーベイに役に立つ文と考えられる．本セグメントからはそのような文を重要文として抽出することを試みる．また「序論」に書かれている重要文の多くは，論文のアブ

[3]後述するように，「序論」のセグメントからの重要文抽出は機械学習に基づく手法を採用したため，この表に示した「本論文の目的は」という手がかり句は実際には用いていない．

ストラクトでも書かれていると思われる．そこで，論文のアブストラクトに書かれている文と同じ文，もしくは内容が似ている文を重要文として抽出する．

表 3.3 は西川らの論文「自動要約における誤り分析の枠組」におけるアブストラクトに出現する文を，表 3.4 は同論文における序論のセグメントに出現する文の一部を示している [4]．2 つの表の各文には文 ID が割り当てられている．「類似文」の列は，アブストラクトと序論に出現する文のうち，互いに似ている文の ID を表わす．表 3.3 における「類似文」の番号は，アブストラクト内のその文に対し，それと似ている文の表 3.4 における ID を表わす．同様に，表 3.4 における「類似文」の番号は，序論内のその文に対し，それと似ている文の表 3.3 における ID を表わす．例えば，アブストラクトに出現する文 1 と序論に出現する文 12 は似ている．この論文では，アブストラクトには 5 つ，序論には 22 の文があるが，それらのうち 4 つは互いに似ていることになる．また，これらの文は，論文の目的や貢献について述べていることがわかる．いくつかの論文について検証した結果，このように「序論」のセグメントと論文のアブストラクトで現れる類似文は，論文の目的や貢献を説明することが多いことがわかった．したがって「序論」のセグメントからはこのような文を重要文として抽出する．

[4]表 3.4 の文 19, 20 における [REF...] は，図表や式，他の節などを参照する LaTeX コマンド \ref にマッチすることを意味する．

表 3.3: アブストラクトの文の例

ID	類似文	文
文 1	12	本稿では自動要約システムの誤り分析の枠組みを提案する.
文 2	15	この誤り分析の枠組みは，要約が満たすべき 3 つの要件と誤った要約が生じる 5 つの原因からなり，要約の誤りをこれらからなる 15 種類の組み合わせに分類する.
文 3		また，システム要約において 15 種類の誤りのうちどの誤りが生じているかを調査する方法もあわせて提案する.
文 4	19	提案する誤り分析の枠組みに基づき，本稿ではまず，システム要約を分析した結果を報告する.
文 5	20	さらに，分析の結果に基づいて要約システムを改良し，誤り分析の結果として得られる知見を用いてシステムを改良することでシステム要約の品質が改善されることを示す.

表 3.4: 「序論」のセグメントの文の例

ID	類似文	文
		（前略）
文 12	1	この状況を鑑み，本稿では，自動要約における誤り分析の枠組みを提案する.
文 13		まず，要約システムが作成する要約が満たすべき 3 つの要件を提案する.
文 14		また，要約システムがこれらの要件を満たせない原因を 5 つ提案する.
文 15	2	3 つの要件と 5 つの原因から，15 種類の具体的な誤りが定義され，本稿では，自動要約における誤りはこれらのいずれかに分類される.
		（中略）
文 19	4	[REF_sc:分析の実践] 節では実際の要約例に含まれる誤りを提案した枠組みに基づいて分析した結果を示す.
文 20	5	[REF_sc:分析に基づく要約システムの改良] 節では [REF_sc:分析の実践] 節で得られた分析の結果に基づいて要約システムを改良し，要約の品質が改善することを示す.
		（後略）

上記のことを実現するため，教師あり機械学習の手法を用いる.「序論」のセグメントに含まれる文のうち，論文のアブスラクトにも出現する文 (もしくは類似した文がアブス

トラクトにも出現する文) は重要文，それ以外は非重要文とする．これにより，重要文と非重要文から構成される文の集合を作成する．これを訓練データとして，文が重要文か否かを判定する二値分類器を学習する．

訓練データの作成は以下のような手続きで行う．論文から「序論」のセグメントを抽出し，それに含まれる文を s_i とおく．s_i が式 (3.3) の条件を満たすとき，その文を重要文と判定し，そうでないときは非重要文と判定してタグ付けをする．

$$\max_{s_a \in A} sim(s_i, s_a) > T \tag{3.3}$$

A は論文のアブストラクトに現れた文の集合であり，s_a はこの集合の要素となる文である．$sim(s_i, s_a)$ は文間の類似度であり，これが閾値 T より大きい文が A に存在するとき，文 s_i と同じ内容の文がアブストラクトに出現しているとみなし，抽出すべき重要文であると判定する．本論文では閾値 $T = 6$ と設定した．文間の類似度は，式 (3.4) のように定義する．

$$sim(s_i, s_a) = \sum_{x \in TG(s_i), y \in TG(s_a)} \delta(x, y) \tag{3.4}$$

TG は文中に含まれる単語 3-gram の集合であり，$\delta(x, y)$ はクロネッカーのデルタ（式 (3.5)）である．

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{3.5}$$

つまり，2 つの文に共通して現われる単語 3-gram の数を類似度と定義する．以下に，表 3.3 の文 1，表 3.4 の文 12，またそれぞれの文から抽出される単語 3-gram の集合と，これらを用いた文間の類似度の計算例を示す．

- 文 1: 本稿では自動要約システムの誤り分析の枠組みを提案する．

  → { 本稿-で-は, で-は-自動要約システム, は-自動要約システム-の,
  自動要約システム-の-誤り分析, の-誤り分析-の, 誤り分析-の-枠組み, の-枠組み-を,
  枠組み-を-提案, を-提案-する，提案-する-(文末) }

- 文 12: この状況を鑑み，本稿では，自動要約における誤り分析の枠組みを提案する．

  → { この-状況-を, 状況-を-鑑み, を-鑑み-本稿, 鑑み-本稿-で, 本稿-で-は,
  で-は-自動要約, は-自動要約-に, 自動要約-に-おける, に-おける-誤り分析,
  おける-誤り分析-の, 誤り分析-の-枠組み, の-枠組み-を, 枠組み-を-提案,
  を-提案-する，提案-する-(文末) }

これらの単語 3-gram の集合には，以下の 6 つが共通して出現する．

本稿-で-は, 誤り分析-の-枠組み, の-枠組み-を, 枠組み-を-提案, を-提案-する
提案-する-(文末)

したがって，式 (3.4) によるこれらの文間の類似度は 6 である．

次に，与えられた文が重要文であるか否かを判定するモデルの学習について述べる．学習アルゴリズムとしては Support Vector Machine(SVM) を用いる．SVM の学習には LIBSVM[5] を用いる．カーネルは線形カーネルを使用し，学習パラメタはデフォルト値とする．

SVM を学習するためには，訓練データの文を素性ベクトルに変換する必要がある．本研究では，素性として文中に含まれる単語の n-gram(n=1,2,3) を使用している．ただし，$n = 1$ のときは自立語（名詞，形容詞，動詞，副詞）のみを素性とし，$n = 2, n = 3$ のときは自立語と付属語を区別せず，全ての単語の並びを素性とする．素性の重みは，単語 n-gram が文中に出現すれば 1，そうでない場合は 0 とする．

機械学習では，訓練データの量が少ないときに，学習に用いる素性数が多すぎると，過学習を起こすことが知られている．特に，単語 3-gram は，3 つの単語の組み合わせとなるため，素性数が非常に多くなる．そのため，簡単な素性選択を行い，有用でないと考えられる素性を削除する．ここでは，訓練データにおける出現頻度が 1 の素性を削除する．

重要文を抽出する「序論」のセグメントが与えられたとき，それに含まれるそれぞれの文に対し，学習した SVM を用いて，それが重要文に該当するか否かを判定する．訓練データを作成したときと同様に，「序論」のセグメントの文から単語の n-gram を抽出し，素性ベクトルを作成する．ただし，訓練データに出現しない素性 (単語 n-gram) は素性ベクトルの作成に使用しない．得られた素性ベクトルを入力とし，SVM で重要文か否かの判定を行う．重要文と判定された全ての文を「序論」のセグメントの要約として出力する．

3.3.2 「関連研究」からの重要文抽出

このセグメントには，しばしば，先行研究と当該論文の研究との差異を説明した文や，先行研究の問題点を指摘しつつその論文の提案手法の特色を強調した文が存在する．このような文を包括的要約に含めると，ユーザはサーベイの際に論文の特徴を知ることができるため，重要文として抽出することをここでの目標とする．

重要文抽出は，この節の冒頭で述べたような標準的な手法を用いる．すなわち，セグメント内の各文について，その重要度 (スコア) を算出し，そのスコアの大きい上位の文を重要文として抽出する．以下，文の重要度を計算する方法について説明する．

先行研究との差異を説明した文や，先行研究の問題点を指摘した文は，典型的な言い回しがあると考えられる．3.2.2 項で述べた「関連研究」のセグメントを抽出する手法では，図 3.3 に示した手がかり句を用い，その手がかり句を含む段落をセグメントとして取り出していた．また，図 3.3 の P1 は論文の特徴を述べる文，P2 は先行研究との比較を述

[5] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

べている文，P3 は先行研究の問題点を指摘する文にマッチすることを狙っていた．したがって，P1, P2,P3 のパターンにマッチする文は，重要文である可能性が高いと考えられる．そのため，「関連研究」のセグメントの文に対して重要度を計算するときは，パターン P1,P2,P3 にマッチするかを考慮する．

「関連研究」のセグメントに出現する文は常に P1,P2,P3 のパターンにマッチするわけではない．しかし，任意の要約率に対する要約を生成する際には，パターンにマッチしない文も含めて，全ての文に対して重要度を計算する必要がある．そのため，本研究では，標準的な単語の TF·IDF に基づく重要文抽出手法も併用する．以上の考え方に基づき，文 s_i の重要度 (スコア) を式 (3.6) のように定義する．

$$Score(s_i) = Score_{tfidf}(s_i) + Score_{par}(P) \tag{3.6}$$

$Score_{tfidf}(s_i)$ は文中に含まれる単語の TF·IDF の値によって決まるスコアで，この値は式 (3.7) で定義される．

$$Score_{tfidf}(s_i) = \sigma(\sum_{w \in s_i} \mathrm{TF} \cdot \mathrm{IDF}(w)\ ) \tag{3.7}$$

w は文 s_i に含まれる単語であり，TF·IDF(w) は w の TF·IDF の値である．本研究における TF·IDF スコアの計算式を式 (3.8) に示す．

$$\mathrm{TF} \cdot \mathrm{IDF}(w) = tf_w^d \cdot idf_w = tf_w^d \cdot \log \frac{N}{df_w} \tag{3.8}$$

tf_w^d は文書 d における単語 w の出現頻度であり，本研究では出現頻度そのもの (raw frequency) を重みとして使用する．idf_w は，複数の文書に現れる単語の重要度を下げるための値で，この値は，単語 w を含む文書の数を表す df_w と，文書の総数を表す N で計算される．文書の総数とは，TF・IDF のスコアを算出する際に用いる論文コーパスに含まれる文書 (論文) の総数である．また本研究では，単語頻度を計算するために日本語形態素解析ツール MeCab(バージョン 0.996) を用いて，文を単語に分割する．この際，複合名詞は一つの単語として扱われるように処理する．

一方，$Score_{par}(P)$ は，s_i を含む段落 P に応じて与えられるスコアであり，式 (3.9) のように定義する．

$$Score_{par}(P) = \sigma(\sum_{s_j \in P} pattern(s_j)\ ) + 1 \tag{3.9}$$

$pat(s_j)$ は，図 3.3 のパターンに 応じて与えられるスコアであり，段落中の文 s_j が P1, P2, P3 にマッチしたとき，それぞれ 10, 3, 2 点とする．

σ は bipolar sigmoid function であり，式 (3.10) のように定義されている．

$$\sigma(x) = \frac{2}{1+e^{-x}} - 1 \tag{3.10}$$

この関数の値域は [0,1] である．すなわち，この関数は，$Score_{tfidf}(s_i)$ を $[0,1]$，$Score_{par}(P)$ を $[1,2]$ の範囲の値に変換するために用いている．以上をまとめると，式 (3.6) は，関連研究に関する内容であると判定された段落に含まれる文を優先的に選択することを意味する．また，関連研究の内容に関連する段落内の文，もしくはそれ以外の段落の文の重要度の優劣は，文中の単語の TF·IDF 値の和によって決めることを意味する．

「関連研究」のセグメントは，3.2.1 項で説明した節のタイトルのマッチングによって抽出する手法，もしくは 3.2.2 項で説明した手がかり句のパターンマッチによって抽出する手法のいずれかによって取得される．重要文の抽出は，セグメントがどちらの手法で抽出されたかによらず，本項で説明した手法で行うことに注意していただきたい．

3.3.3 「提案手法」からの重要文抽出

本項では，「提案手法」のセグメントからの重要文抽出の構想について述べる．「提案手法」のセグメントからは，その論文の提案手法の概略を説明する文を重要文として抽出することを考えている．多くの場合，手法の概要は節や項 (subsection) の先頭に書かれることが多いため，重要文抽出の際には文の位置が有効な手がかりになると思われる．また，提案手法の処理の流れは図で提示されることがしばしばある．そこで，重要文を抽出するとともに，提案手法の流れや概略などを表す図も抽出し，要約に含めることで，サーベイの際に提案手法の概略が速やかに把握できるようになると思われる．

3.3.4 「評価実験」からの重要文抽出

本項では，「評価実験」のセグメントからの重要文抽出の構想について述べる．「評価実験」のセグメントからは，実験の設定や実験の結果を説明している文を重要文として抽出することを考えている．実験の設定についての説明は，評価実験の節の冒頭に書かれることが多い．また，実験の結果は表やグラフを用いて表わされることが多い．表やグラフは実験結果を把握しやすいため，これらを検出して重要文と共に抽出することは有望である．また，実験結果の表やグラフが複数ある場合は，最も主要な結果を示すものを選択する処理も必要となる．

3.4 重要文の結合

包括的要約生成の最後のステップは，各セグメントから抽出された重要文を統合することである．このとき，各セグメントから抽出した重要文は元の論文での出現順に結合す

ることにする．また，セグメント毎に重要文をまとめてから，表形式で提示することで，ユーザが異なる観点から論文の要点を把握できるようにする．

要約率のコントロールは重要な課題である．要約率を満たすように重要文を選択する際，各セグメント毎に要約率を満たすように選択するのか，あるいは「序論」のような重要と考えられるセグメントからより多くの重要文を選択するのかは，今後検討する必要がある．

3.5 データ構造

ここでは，提案手法を実装したシステムにおいて，要約生成のために論文を解析した結果を格納するためのデータ構造について述べる．本章で述べた手法で解析するセグメント構造及び重要文抽出のための文のスコアの情報は，図 3.6 のような階層構造で表わされる．以下，その詳細を説明する．

- リファレンス structure が指す配列の最初のインデックス（文）は，論文の全ての文の数に相当する大きさの配列を指す．この配列の下位の構造の sentence は論文中の文そのもの，rel score は式 (3.6) によって算出された「関連研究」のセグメントから重要文を抽出するための文のスコア，morpheme は文に現れた単語の文書内における頻度，local tf score は文に現れた単語の頻度の和，tf idf score は式 (3.7) による文中の単語の TF·IDF スコアの和を表す．

- リファレンス structure が指す配列の 2 番目のインデックス（概要）は，著者によって書かれた論文のアブストラクトに関する情報を含む．

- リファレンス structure が指す配列の 3 番目のインデックス（序論）からは，各セグメントの情報を保存する．図 3.6 では，それぞれのセグメントに一つのインデックスが割り当てられているが，2 つ以上の節が同じセグメントに分類される場合は，その数の分だけインデックスが増える．下位のデータ構造の type にはセグメントの種類，title には節や項のタイトル，subsection には項に関する情報，start position, end position, section end にはそれぞれ節の最初の文のインデックス，節 (section) のタイトルとその節の下位に位置する最初の項 (subsection) のタイトルの間に書かれたテキストの最後の文のインデックス，節の最後の文のインデックスが入る．paragraphs には，節の中の各段落の最初の文のインデックスが入る．

- リファレンス structure が指す配列の最後のインデックス（関連研究段落）が指すデータ構造は，「関連研究」のセグメントが 3.2.1 項で提案した手法によって検出できなかった場合，3.2.2 項で抽出した段落の情報を保存するためのものである．このインデックスは，3.2.1 項の手法で「関連研究」セグメントが抽出できた場合には生成されない．

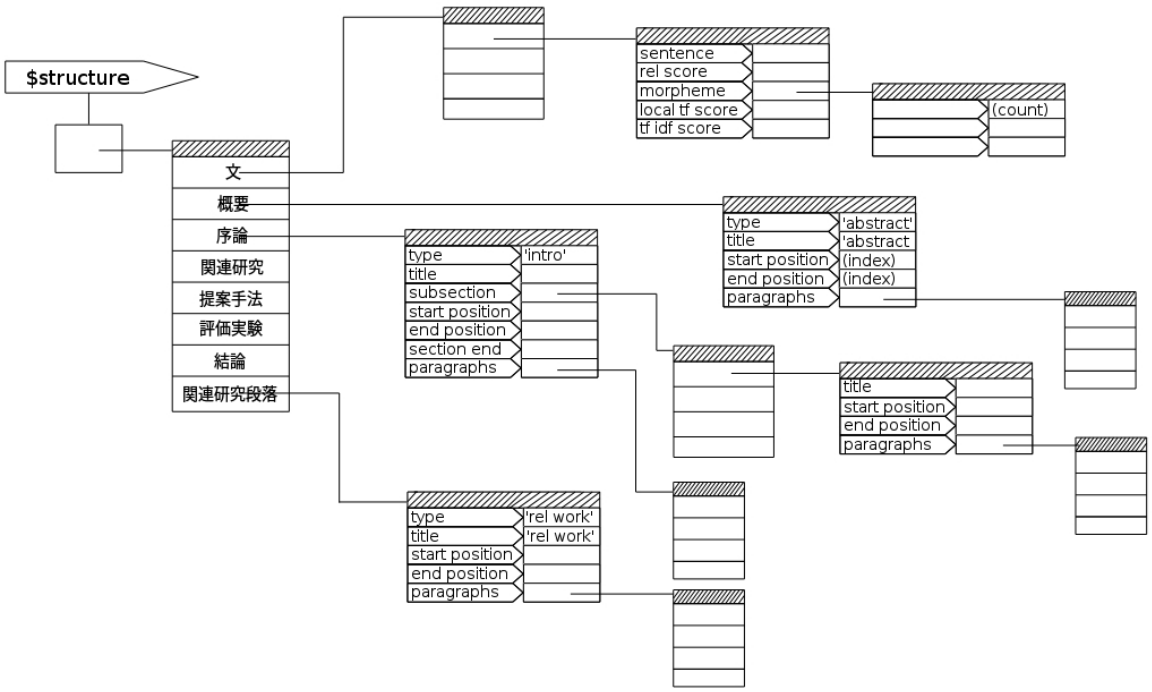

図 3.6: セグメント構造の解析結果の一部

第4章　評価実験

4.1　実験データ

本研究では，提案システムの実装や評価に用いるデータとして，言語処理学会論文誌 LaTeX コーパス [1] を使用する．このコーパスは，会誌「自然言語処理」に掲載された論文の LaTeX のソースファイルを集めたデータ集である．本研究では，日本語を対象とした自動要約の手法を研究するため，同コーパスにおける日本語で書かれた論文のみを使用する．実験では，ランダムに選択した 30 件の論文の LaTeX ファイルをテストデータとし，388 件の論文を訓練・開発データとしている．この訓練・開発データは，3.2.1 項で述べた「序論」のセグメントから重要文を選択する二値分類器の学習に用いるほか，提案手法を設計する際にも参照した．例えば，表 3.1 のキーワードの選定，図 3.3 のパターンの作成，式 (3.8) の TF・IDF のスコアの算出は，この 388 件の訓練・開発データの論文を参照して行った．

4.2　セグメント構造解析の評価

まず，節のタイトルを手がかりとしてセグメントを決定する手法を評価する．この手法では，キーワードリストを作成し，節のタイトルに対するパターンマッチングを行うことで，テストデータ 30 件の論文におけるそれぞれの節を「序論」「関連研究」「提案手法」「実験結果」「結論」のセグメントのいずれかに分類する．

まず，セグメントの構造解析を精度と再現率で評価する．精度と再現率の定義をそれぞれ式（4.1）と式（4.2）に示す．

$$\text{精度} = \frac{\text{抽出された正しいセグメント（節）の数}}{\text{抽出されたセグメント（節）の数}} \tag{4.1}$$

$$\text{再現率} = \frac{\text{抽出された正しいセグメント（節）の数}}{\text{論文中のセグメント（節）の数}} \tag{4.2}$$

セグメントごとの精度と再現率を表 4.1 に示す．

[1] http://www.anlp.jp/resource/journal_latex/index.html

表 4.1: セグメント分割の結果

セグメント	序論	関連研究	提案手法	評価実験	結論
精度	1.0	1.0	0.83	1.0	1.0
再現率	1.0	0.62	1.0	0.73	0.91

精度は 5 つのセグメントのいずれも高い．本手法では，「提案手法」以外のセグメントは節のタイトルに対するキーワードのパターンマッチで抽出し，一方「提案手法」のセグメントはパターンにマッチしない節を全て選択している．「提案手法」以外のセグメントはキーワードのマッチングにより抽出しているが，これらのセグメントの精度は 100%であるのに対し，「提案手法」のセグメントの精度は 83%とやや劣る．一方，再現率は「関連研究」と「評価実験」のセグメントではやや低いが，それ以外は十分に高い．「関連研究」については，節のタイトルを見るだけでは関連研究に関する内容が書かれている節であるのか判定できない論文（図 4.1）や，「関連研究」の内容が独立した節ではなくある節内の一部の項に含まれている論文（図 4.2）があったのが原因である．図 4.1 の論文は日本語の係り受け解析に関する論文であり，「日本語係り受け解析」というタイトルの節に関連研究の説明があるが，表 3.1 に示したキーワードでは検出できなかった．なお，この図の \section は節を区切る際の LaTeX コマンドを意味する．図 4.2 の例では，「関連研究」が項 (subsection) のタイトルになっているため，検出できなかった．この図の \subsection は項を区切る際の LaTeX コマンドである．

```
文書 ID   : V18N04-02
タイトル : shWiiFit Reduce Dependency Parsing
著者      : 浅原 正幸
...（前略）...
\section{日本語係り受け解析}
...（後略）...
```

図 4.1: 「関連研究」の抽出失敗例：タイトルではマッチできない場合

文書 ID　：V06N05-03
タイトル：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
著者　　：難波 英嗣，奥村 学
...（前略）...
\section{サーベイ論文作成}
...（中略）...
\subection{サーベイ論文作成のポイント}
...（中略）...
\subsection{関連研究}
...（後略）...

図 4.2: 「関連研究」の抽出失敗例：下位の項になっている場合

「評価実験」については，「関連研究」と同じく，節のタイトルを見るだけでは節の中に「評価実験」に関する内容が含まれているか判定できない論文（図 4.3）や，提案手法を論じる節がいくつかの項から構成され，その中の一つの項に評価実験が書かれている論文があり（図 4.4），このような場合には「評価実験」のセグメントの抽出に失敗していた．図 4.3 の例では，「参照用テストセット」というタイトルの節に評価実験に関する説明があるが，表 4.1 のキーワードにマッチしないため，検出できなかった．図 4.4 の例では，「レシピ用語の自動認識」というタイトルの節に提案手法の説明がされているが，その節の中に「未知のレシピ用語タグの推定事例」というタイトルの項があり，この中に評価実験に関する記述があったが，本手法は節を単位にセグメントを検出するので，抽出できなかった．

文書 ID　：V15N03-03
タイトル：情報アクセス対話のための質問応答技術評価タスク
著者　　：加藤 恒昭，福本 淳一，桝井 文人，神門 典子
...（前略）...
\section{参照用テストセット}
...（後略）...

図 4.3: 「評価実験」の抽出失敗例：タイトルではマッチできない場合

文書 ID　：V22N02-02
タイトル：レシピ用語の定義とその自動認識のためのタグ付与コーパスの構築
著者　　：笹田 鉄郎，森 信介，山肩 洋子，前田 浩邦，河原 達也
...（前略）...
\section{レシピ用語の自動認識}
...（中略）...
\subection{レシピ用語の自動認識と精度評価}
...（中略）...
\subsection{未知のレシピ用語タグの推定事例}
...（後略）...

図 4.4: 「評価実験」の抽出失敗例：下位の項になっている場合

表 4.2 は，テストデータとして用いた 30 件の個々の論文に対するセグメント抽出の結果を示している．

表 4.2: 個々の論文に対するセグメント抽出の結果

文書ID	節のタイトルで関連研究抽出	「序論」	「関連研究」	「提案手法」	「評価実験」	「結論」
V01N01-01	0	1:1/1:1/1	0:0/0:0/0	1:2/2:2/2	1:1/1:1/1	1:1/1:1/1
V02N04-03	0	1:1/1:1/1	0:0/0:0/0	1:3/4:3/3	1:1/1:1/1	1:1/1:1/1
V03N03-03	0	1:1/1:1/1	0:0/0:0/1	1:2/4:2/2	1:1/1:1/2	1:1/1:1/1
V04N01-04	0	1:1/1:1/1	0:0/0:0/0	1:2/2:2/2	1:1/1:1/1	0:0/0:0/1
V04N04-01	0	1:1/1:1/1	0:0/0:0/0	1:3/3:3/3	1:1/1:1/1	1:1/1:1/1
V06N02-06	0	1:1/1:1/1	0:0/0:0/1	1:2/3:2/2	0:0/0:0/1	1:1/1:1/1
V06N05-03	0	1:1/1:1/1	0:0/0:0/1	1:3/4:3/3	1:1/1:1/1	1:1/1:1/1
V07N02-07	0	1:1/1:1/1	0:0/0:0/0	1:1/1:1/1	1:2/2:2/2	1:1/1:1/1
V07N04-07	0	1:1/1:1/1	0:0/0:0/0	1:8/8:8/8	1:1/1:1/1	1:1/1:1/1
V08N04-02	0	1:1/1:1/1	0:0/0:0/0	1:2/2:2/2	1:2/2:2/2	1:1/1:1/1
V09N04-04	1	1:1/1:1/1	1:1/1:1/1	1:3/3:3/3	1:1/1:1/1	1:1/1:1/1
V10N01-04	0	1:1/1:1/1	0:0/0:0/0	1:2/4:2/2	1:1/1:1/3	1:2/2:2/2
V10N05-02	0	1:1/1:1/1	0:0/0:0/1	1:3/4:3/3	0:0/0:0/1	1:1/1:1/1
V11N05-07	1	1:1/1:1/1	1:1/1:1/1	1:2/2:2/2	1:1/1:1/1	1:1/1:1/1
V12N05-05	0	1:1/1:1/1	0:0/0:0/0	1:5/5:5/5	1:1/1:1/1	1:1/1:1/1
V13N03-06	0	1:1/1:1/1	0:0/0:0/1	1:3/5:3/3	0:0/0:0/1	1:1/1:1/1
V14N02-01	1	1:1/1:1/1	1:1/1:1/1	1:2/3:2/2	1:1/1:1/1	1:1/1:1/2
V14N03-11	0	1:1/1:1/1	0:0/0:0/0	1:2/2:2/2	1:2/2:2/2	1:2/2:2/2
V14N05-05	1	1:1/1:1/1	1:1/1:1/1	1:4/4:4/4	0:0/0:0/1	1:1/1:1/1
V15N03-03	1	1:1/1:1/1	1:1/1:1/1	1:1/2:1/1	1:1/1:1/2	1:1/1:1/1
V16N01-01	0	1:1/1:1/1	0:0/0:0/0	1:2/2:2/2	1:1/1:1/1	1:1/1:1/1
V16N04-04	0	1:1/1:1/1	0:0/0:0/0	1:3/3:3/3	1:2/2:2/2	1:1/1:1/1
V17N01-11	0	1:1/1:1/1	0:0/0:0/0	1:2/3:2/2	0:0/0:0/1	1:1/1:1/1
V17N05-02	1	1:1/1:1/1	1:1/1:1/1	1:2/2:2/2	0:0/0:0/0	1:2/2:2/2
V18N04-02	0	1:1/1:1/1	0:0/0:0/1	1:2/3:3/3	1:1/1:1/1	1:1/1:1/1
V19N05-02	1	1:1/1:1/1	1:1/1:1/1	1:3/3:3/3	1:1/1:1/1	1:1/1:1/1
V20N03-03	1	1:1/1:1/1	1:1/1:1/1	1:2/2:2/2	1:1/1:1/1	1:1/1:1/1
V21N01-03	1	1:1/1:1/1	1:1/1:1/1	1:2/2:2/2	1:1/1:1/1	1:1/1:1/1
V21N03-03	0	1:1/1:1/1	0:0/0:0/0	1:3/3:3/3	1:1/1:1/1	1:1/1:1/1
V22N02-02	1	1:1/1:1/1	1:1/1:1/1	1:2/2:2/2	0:0/0:0/1	1:1/1:1/2

2 列目「節のタイトルで関連研究抽出」では，節のタイトルを手がかり句とした手法で「関連研究」のセグメントが抽出できた論文に 1 を，できなかった論文に 0 を付けている．3 列目以降は各セグメントについての結果を a:b/c:d/e という形式で示している．a はセグメントの抽出に成功したかを表わす (成功したとき 1，失敗したとき 0)．b と c は式 (4.1) に示した精度の分子と分母を表わす．d と e は式 (4.2) に示した再現率の分子と分母を表

わす．ただし，「節のタイトルで関連研究抽出」の列が 0 になっている論文では，a, b, c, d は常に 0 である．4 列目の結果を見ると，a=0 かつ e=1 となっている論文，すなわち関連研究について述べている節があるのにも関わらず，節のタイトルを手がかりとする手法でそれを抽出できなかった論文が多いことがわかる．このことが，表 4.1 に示したように，「関連研究」のセグメントの再現率が低い原因となっている．

次に，それぞれのセグメントの抽出率を表 4.3 (a) に示す．抽出率の定義は式 (4.3) の通りである．

表 4.3: セグメントの抽出率

(a) 節のタイトルを手がかりとする手法のみ

セグメント	序論	関連研究	提案手法	評価実験	結論
抽出率	100%	33%	100%	80%	96%

(b) 関連研究の手がかり句に基づく手法を併用したとき

セグメント	序論	関連研究	提案手法	評価実験	結論
抽出率	100%	100%	100%	80%	96%

$$\text{抽出率} = \frac{\text{セグメントを検出できた論文の数}}{\text{論文の総数}} \tag{4.3}$$

すなわち，抽出率は，正解・不正解に関わらずセグメントを抽出できた論文の割合である．セグメントの抽出率は「関連研究」を除いて高い．「関連研究」については，30 件中 10 件の論文しかセグメントを抽出できなかった．これは，既に述べたように，「関連研究」の内容が独立した節ではなく，節の中の一部の段落として含まれている論文が多かったためである．

「関連研究」のセグメントが抽出できなかった 20 件の論文については，論文全体を段落に分割した上で 3.2.2 項で述べた関連研究の手がかり句のパタンマッチによる手法で，関連研究について述べているセグメント（段落）を取り出すことを試みる．このときの抽出率を表 4.3(b) に示す．「関連研究」についてはテストデータの全ての論文からセグメントを抽出できるようになった．それ以外のセグメントの抽出率は変化しない．

次に，抽出されたセグメント (段落) が関連研究に関する記述を含んでいる場合には正解とみなし，その精度を算出する．正解の判定は人手で行う．精度の定義を式 (4.4) に示す．

$$\text{精度} = \frac{\text{抽出されたセグメント (段落) のうち正解とみなせるものの数}}{\text{抽出されたセグメント (段落) の総数}} \tag{4.4}$$

その結果，精度は 65%となった．詳細を表 4.4 に示す．

図 4.5 は関連研究について述べながら論文の特徴を説明している段落を抽出した例である．下線は手がかり句にマッチした箇所を表わす．抽出した段落は「関連研究」のセグメント（段落）とみなせる．

表 4.4: 関連研究の手がかり句に基づくセグメント抽出の精度

抽出セグメント数	正解セグメント数	精度
20	13	0.65

文書 ID　：V16N01-01
タイトル：話し言葉における引用節・挿入節の自動認定および係り受け解析への応用
著者　　：浜辺 良二，内元 清貴，河原 達也，井佐原 均

CSJ では，挿入節の終端の文節に「挿入節」というラベルが付与されている．
挿入節の終端は基本的に強境界となっているが，挿入節を越えて前方から後方に
係る係り受けが存在するため，文境界ではなく挿入節の終端と認定される．

従来 研究 では，話し言葉において節境界の曖昧さが係り受け解析に及ぼす影響
については，ほとんど考慮されていなかった．
下岡ら [CITE] は，話し言葉では文境界が曖昧であることが係り受け解析に与える
影響が最も大きいことを指摘し，その影響を定量的に示した．
彼らは，正しい文境界の情報を与えることにより，文境界を自動推定した場合に
比べて約 3%高い係り受け解析精度が得られると報告している．
また，文境界を推定する方法および文境界の自動推定結果を係り受け解析に利用
する 方法 を 提案 し，その有効性も示した．
しかし，その他の節境界については，係り受け解析に及ぼす影響は
明らかではなかった．
...（中略）...
逆に，引用節・挿入節の範囲を取得することができれば，係り受け解析精度の
向上が期待 できる が，そこまでは明らかにはされて いない．
そこで，本論文 では，引用節・挿入節を自動認定する 手法，および，その結果を
利用して係り受け解析を行なう 手法 を 提案 し，引用節・挿入節を自動認定した
結果を用いることで係り受け解析精度が有意に向上することを示す．
手法 については，[REF_sec:method] 章で詳しく述べる．

図 4.5: 「関連研究」のセグメント (段落) の抽出例

4.3 重要文抽出の評価

4.3.1 「序論」のセグメントからの重要文抽出の評価

まず，3.3.1 項で述べた「序論」のセグメントからの重要文抽出の手法を評価する．テストデータの正解の要約，すなわち正解の重要文のセットは，訓練データと同じ方法で作成する．すなわち，テストデータの「序論」のセグメントに含まれている文が，その論文のアブストラクト内の文と同じ意味を持つと判断したとき，その文を要約に含めるべき重要文であると判定する．重要文抽出の評価には，精度，再現率，F 値を用いた．精度，再現率，F 値の定義をそれぞれ式 (4.5)，(4.6)，(4.7) に示す．評価結果を表 4.5 に示す．精度，再現率，F 値はいずれも 30%程度であり，改善の余地がある．

$$\text{精度}\ (P) = \frac{\text{選択された正解の重要文の数}}{\text{システムが重要文として選択した文の数}} \tag{4.5}$$

$$\text{再現率}\ (R) = \frac{\text{選択された正解の重要文の数}}{\text{正解の重要文の数}} \tag{4.6}$$

$$F\ \text{値} = \frac{2PR}{P+R} \tag{4.7}$$

表 4.5: 「序論」のセグメントからの重要文抽出結果

精度	再現率	F 値
0.31	0.29	0.30

表 4.9 は，笹田らの論文「レシピ用語の定義とその自動認識のためのタグ付与コーパスの構築」 における「はじめに」という節の文の一部である．この節は提案手法によって「序論」のセグメントとして検出されている．この表では，正解としてタグ付けされた文は「正解」の列にチェックを，提案手法によって重要文として選択された文は「選択」の列にチェックを入れている．文 3 は正解文であり，またシステムも重要文として抽出している．正解の文は文 1 と文 3 の 2 つである．文 1 は論文の背景について説明しており，文 3 は論文中の用語 (固有表現) の定義について述べている．これらは「序論」のセグメントから抽出すべき正解文ではない．「序論」のセグメントからは，主に論文の目的や研究の貢献について述べている文を抽出することが目的であった．したがって，文 1 も文 3 も「序論」のセグメントから抽出するべき正解文ではない．不適切な文が正解となっている理由は，正解文が自動的に選択されているからである．これらの文の式（3.3）の左辺の値が閾値 T より大きく，論文のアブストラクトに含まれる文と十分に高い類似度を持っていると判定された．不適切な文を正解の重要文と判定した誤りは，重要文を判定する分類器の訓練データ作成時にも発生していると考えられる．したがって，訓練データや正解

を作成する際，論文の目的と貢献について述べている文を選択できるように手法を改善することが必要とされる．一方，システムが重要文として抽出したのは文 3 と文 18 である．既に説明したように，文 3 は抽出するべき重要文ではないが，文 18 は論文の目的を述べており，重要文と言える．ただし，この文は正解ではないので，精度や再現率を算出する際には誤りと扱われている．また，文 24 は研究の貢献を述べている文であるが，提案手法によって抽出されていない．これは，重要文抽出のための素性として文中に含まれる単語の n-gram(n=1,2,3) だけしか用いていないことが原因のひとつと考えられる．「序論」のセグメントの重要文に現われる言語的な特徴は何かを探究し，これを学習素性として分類モデルに組み込むことが今後の課題である．

表 4.6: 「序論」のセグメントからの重要文抽出例

ID	正解	選択	文
文 1	✓		自然言語処理において，単語認識（形態素解析や品詞推定など）の次に実用化可能な課題は，用語の抽出であろう．
文 2			この用語の定義としてよく知られているのは，人名や組織名，あるいは金額などを含む固有表現である．
文 3	✓	✓	固有表現は，単語列とその種類の組であり，新聞等に記述される内容に対する検索等のために 7 種類（後に 8 種類となる）が定義されている [CITE]．
			（中略）
文 14			本論文では，この過程の実例を示し，ある固有表現の定義の下である程度高い精度の自動認識器を手早く構築するための知見について述べる．
文 15			本論文で述べる固有表現は，以下の条件を満たすとする．
文 16			以上の条件は，品詞タグ付けに代表される単語を単位としたタグ付けの手法を容易に適用させるためのものである．
文 17			その一方で，日本語や中国語のように単語分かち書きの必要な言語に対しては，あらかじめ単語分割のプロセスを経る必要があるという問題も生じるが，本論文では単語分割を議論の対象としないものとする．
文 18		✓	本論文では，題材を料理のレシピとし，さまざまな応用に重要と考えられる単語列を定義し，ある程度実用的な精度の自動認識を実現する方法について述べる．
文 19			例えば，「フライ 返し」という単語列には「フライ」という食材を表す単語が含まれるが，一般的に「フライ返し」は道具であり，「フライ 返し」という単語列全体を道具として自動認識する必要がある．
文 20			本論文ではこれらの単語列をレシピ用語と定義してタグ付与コーパスの構築を行い，上述した固有表現認識の手法に基づく自動認識を目指す．
			（中略）
文 24			本論文で対象とするレシピテキストはユーザ生成コンテンツ (User Generated Contents; UGC) であり，そのようなデータを対象とした実際のタグ定義ならびにアノテーション作業についての知見やレシピ用語の自動認識実験から得られた知見は，ネット上への書き込みに対する分析など様々な今日的な課題の解決の際に参考になると考えられる．

4.3.2 「関連研究」のセグメントからの重要文抽出の評価

続いて 3.3.2 項で述べた「関連研究」のセグメントからの重要文の抽出手法を評価する．テストデータの正解の要約は，「関連研究」のセグメントとして抽出された節や段落から重要文と思われる文を人手で選択して決める．選択の基準としては，関連する研究に対する当該論文の位置付けを述べている箇所の中で，当該論文の特徴を説明している文を要約に含めるべき重要文とする．この際，正解とする重要文の数には制限を設けていない．また，「関連研究」のセグメント以外の部分からも，例えば「序論」や「結論」のセグメントに関連研究について言及しながら論文の特徴を強調するような文があれば，それも正解の重要文として選択する．

本研究における「関連研究」のセグメントからの重要文抽出手法では，文のスコアを計算し，その上位 N 個の文を重要文として選択する．今回の実験では $N = 4$ と設定する．重要文抽出の精度，再現率，F 値を表 4.7 に示す．

表 4.7: 「関連研究」のセグメントからの重要文抽出結果

	精度	再現率	F 値
全体 (30 論文)	0.21	0.24	0.22
タイトル (10 論文)	0.20	0.32	0.25
手がかり句 (20 論文)	0.21	0.22	0.22

2 行目の「全体」はテストデータ全体の 30 論文に対する評価結果，3 行目の「タイトル」は節のタイトルに対するパタンマッチによってセグメント抽出できた 10 論文に対する評価結果，4 行目の「手がかり句」は手がかり句のパタンマッチによってセグメント（段落）を抽出した 20 論文に対する評価結果を示す．

精度は，タイトルのパタンマッチで検出されたセグメント，手がかり句のパタンマッチで検出されたセグメントのいずれも 20%程度であった．一方，再現率は，前者のセグメントが 32%，後者のセグメントが 22%で，前者の方が 10%程度高かった．これは，タイトルのパターンマッチングによってセグメント分割を行う方が，手がかり句のパターンマッチングによってセグメントを抽出する方法に比べてセグメント抽出の精度が高いことが原因であると思われる．表 4.1 で示したように節のタイトルのパターンマッチによってセグメントを検出する手法の精度は 100%，表 4.4 で示したように関連研究の手がかり句によってセグメントを検出する手法の精度は 65%である．後者の手法で抽出されるセグメントは節単位ではなく段落単位であるが，関連研究とは関係のない段落が含まれていることも多く，そのような段落から抽出された重要文はほとんど不正解となった．

表 4.8 は，関連研究に関する手がかり句により段落をセグメントとして抽出する手法を用いた 20 件の論文に対する重要文抽出結果の詳細を示している．「論文 ID」は論文の識別番号，「重要文数」は論文中に出現する正解の重要文の数，「セグメント判定」は抽出されたセグメントが関連研究に関する内容であるかの判定 (1 はそうであるとき，0 はそう

でないとき),「セグメント内の重要文数」はセグメント内に出現する正解の重要文の数,「抽出された正解重要文数」は本手法によって選択された重要文のうち正解の数を表わす．20 論文中 7 つの論文で「セグメント判定」が 0 となっており，関連研究に関して論じた段落の抽出に失敗している．これらの論文からは重要文をひとつも抽出できていない．また,「セグメント判定」が 1 となっている 13 個の論文について,「重要文数」の和は 52 であるのに対し,「セグメント内の重要文数」の和は 30 であり，検出されたセグメントの中には 58%の重要文しか含まれていない．これらは，表 4.7 において，手がかり句によって「関連研究」のセグメントを抽出した論文における重要文抽出の再現率が低い事実を裏づけている．すなわち，セグメント抽出の段階で，抽出するべき多くの重要文を取り出せていない．一方,「抽出された正解重要文数」の和は 17 であり，抽出されたセグメント内に存在する重要文の 57%に相当する．これは表 4.7 に示した再現率 0.22 よりもかなり高い．以上から，重要文抽出の再現率が低い主な要因はセグメント抽出の誤りであると言える．

表 4.8: 関連研究の手がかり句によってセグメントを検出した 20 論文に対する重要文抽出の評価

文書ID	重要文数	セグメント判定	セグメント内の重要文数	抽出された正解重要文数
V01N01-01	4	0	0	0
V02N04-03	3	1	2	0
V03N03-03	5	0	0	0
V04N01-04	2	1	1	0
V04N04-01	4	1	1	1
V06N02-06	5	1	1	1
V06N05-03	4	0	0	0
V07N02-07	5	1	5	2
V07N04-07	1	0	0	0
V08N04-02	10	1	5	3
V10N01-04	5	1	5	3
V10N05-02	3	1	3	2
V12N05-05	2	1	2	2
V13N03-06	1	1	0	0
V14N03-11	1	1	0	0
V16N01-01	6	1	1	1
V16N04-04	5	1	4	2
V17N01-11	7	0	0	0
V18N04-02	3	0	0	0
V21N03-03	2	0	0	0

図 4.9 は，内山らの論文「統計的手法による分野非依存のテキスト分割」（文書 ID は V08N04-02）における「関連研究」のセグメントからの重要文抽出の結果を示している．

「正解」の列は人手で判定した正解の重要文を，「抽出」は提案手法によって抽出された文を表わす．また，セグメントの全ての文ではなく，一部の文を抜粋して表示している．

表 4.9: 「関連研究」のセグメントからの重要文抽出の例

ID	正解	抽出	文
文 21	✓		本稿で述べる手法も，これらの従来手法と同様に，訓練データを利用せずに，テキスト内の単語分布のみを利用してテキストを分割する．
文 24	✓		本稿で述べる手法は，テキストの分割確率が最大となるような分割を選択するというものである．
文 31	✓		しかし，そのような方法は，訓練データが利用できない分野については適用できないので，我々の目的である，テキスト分割の結果を利用して，長い文書を要約したり，講演のディクテーション結果を要約するためのテキスト分割手法としては適さない．
文 199		✓	これから分かるように，最小コスト解よりも粒度の細かい分割が必要なときには，再帰的分割をした方が精度良く分割ができる．
文 201	✓		提案手法は，分割確率最大化という観点からテキスト分割を定式化した．
文 202	✓	✓	これに類似の手法として，訓練データを利用したテキスト分割では，[CITE] が隠れマルコフモデルに基づいて，複数ニュースを個々のニュースに分割しているが，訓練データを利用しないテキスト分割では，類似の研究はない．
文 204	✓		そのため，彼等のアプローチでは，たとえば，トピックの平均の長さなどを直接取り込むことが難しい．
文 205	✓		一方，我々のアプローチでは，このことは素直に表現できる．
文 206	✓	✓	たとえば，[CITE] と同様に，トピックの長さ [MATH] が，平均長 [MATH]，標準偏差 [MATH] の正規分布 [MATH] に従うと仮定すると，単純な拡張としては，([REF_eq:cS_i]) 式を，[MATH] として，以下のようにすれば，トピックの長さが平均と同じくなるような分割が優先される．
文 207	✓	✓	更に，彼等の手法と我々の手法との大きな違いは，彼等が単語の確率を訓練データから推定しているのに対して，我々は，単語の確率を分割対象のテキストから推定している点である．
文 215	✓		そのために，我々は，本稿では，大域的な最小コスト解よりも細かい分割が必要な場合には，再帰的な分割を適用し，それは有効ではあったが，より有効な分割方法を考えることは今後の課題としたい．

この例では，抽出した 4 つの文のうち，文 202, 206, 207 が正解文であり，精度は 75%と高い．一方，正解文の数は 10 なので，再現率は 33%に留まっている．正解文をみると，前後の文脈を読まないと重要文であるか判断できないものも含まれている．しかし，提案手法では，個々の文のスコアは互いに独立に計算されている．スコアの計算に文脈も考慮する必要があるだろう．また，今回の実験では，「関連研究」のセグメントから 4 つの重要文を抽出したが，この例のように正解文の数が 4 つ以上になっている論文が多く，このことが再現率を低くする原因のひとつとなっていることが分かった．

第5章　おわりに

5.1　まとめ

本研究では，サーベイの労力を軽減させることを目指し，学術論文の目的，貢献，関連研究との位置付け，提案手法，評価実験など，論文の主要な要点を全て含む要約，「包括的要約」を自動生成することを提唱した．この包括的要約を自動生成するため，まず，多くの学術論文が共通して持つ典型的なセグメント構造に着目し，論文を「序論」「関連研究」「提案手法」「実験結果」「結論」の５つのセグメントに分割する手法を提案した．次に，各セグメントが持つ特徴を考慮して，セグメントの種類ごとに適した重要文選択手法を開発し，この手法をそれぞれのセグメントに適用することで重要文を抽出する手法を提案した．最終的に，それぞれのセグメントの重要文抽出結果を合わせることで包括的要約を生成する．このような学術論文の構造を考慮して要約を生成する研究はこれまで行われていなかった．

本研究は，日本語を対象とした自動要約の手法を研究した．そのため，本研究で提案する自動要約システムの実装や評価に用いるデータとしては，会誌「自然言語処理」に掲載された論文の LaTeX のソースファイルを集めたデータ集である「言語処理学会論文誌 LaTeX コーパス」のうち，日本語で書かれた論文を使用した．これらの論文データの内，実験ではランダムに選択した 30 件の論文の LaTeX ファイルをテストデータとし，388 件の論文を訓練及び開発データとして利用した．この訓練及び開発データは，「序論」のセグメントから重要文を選択する二値分類器の学習に用いるほか，提案手法を設計する際にも参照した．

以下，本論文で提案した手法とその評価をまとめる．まず，元の論文のセグメント構造を解析する．セグメント構造の解析は 2 つの手法によって行う．一つは節のタイトルを手がかりとする手法，もう一つは関連研究の手がかり句に基づく手法である．

節のタイトルを手がかりとする手法では，セグメントの節のタイトルによく使われると思われる表現をキーワードのリストとして用意し，このキーワードが節のタイトルに含まれていれば，その節をセグメントとして抽出した．このとき，「序論」「関連研究」「評価実験」「結論」の４つのセグメントはキーワードのリストを用いる手法で同定し，そのいずれにも該当していない節を「提案手法」のセグメントとみなした．テストデータ 30 件の論文を用いた評価実験では，セグメント抽出の精度は 5 つのセグメントのいずれも高かった．しかし，再現率は「序論」「提案手法」「結論」は充分に高いものの，「関連研究」と「評価実験」のセグメントではやや低かった．一方，それぞれのセグメントの抽出率は

「関連研究」を除いて高かった．

「関連研究」のセグメントについては，節のタイトルを手がかりとする手法だけでは，セグメントが検出されないことが多かった．そのため，節のタイトルを手がかりとする手法で「関連研究」のセグメントが抽出できなかった論文については，論文全体を段落に分割した上で，関連研究の手がかり句に基づく手法で，関連研究について述べている段落の抽出を試みた．「関連研究」のセグメントの中で典型的に使われると考えられる手がかり句のパターンにマッチする文があれば，その段落とその前に出現する 2 つの段落を「関連研究」のセグメントとして抽出した．抽出されたセグメント（段落）が関連研究に関する記述を含んでいる場合には正解とみなし，その精度を算出したところ，65%となった．この際，正解の判定は人手で行った．

次に，分割したそれぞれのセグメントから重要文を抽出する手法を提案した．本論文ではそれぞれのセグメントに適した重要文抽出手法を開発する．また本論文では，論文の目的，貢献，提案手法の概略，主要な評価結果について述べている文を包括的要約に含むべき重要文と定義した．

「序論」のセグメントから抽出すべき重要文は，論文の目的及び貢献について述べている文であると決めた．「序論」に書かれている重要文の多くは，アブストラクトでも書かれていると思われるため，2 つの文に共通して現われる単語 3-gram の数を類似度と定義し，アブストラクトに書かれている文と同じ文，もしくは内容が似ている文を重要文として抽出した．これらの重要文は，与えられた文が重要文であるか否かを判定する二値分類器の学習に用いた．また，システムの性能評価のための正解データも同じ方法で作成した．

与えられた文が重要文であるか否かを判定するモデルの学習には，学習アルゴリズム Support Vector Machine(SVM) を用い，SVM の学習には LIBSVM を用いた．学習素性として文中に含まれる単語の n-gram(n=1,2,3) を使用した．テストデータ 30 件における重要文抽出の評価には，精度，再現率，F 値を用いた．実験の結果，提案手法による「序論」のセグメントからの重要文抽出の精度，再現率，F 値はいずれも 30%程度であった．

「関連研究」のセグメントには，先行研究と当該論文の研究との差異を説明した文や，先行研究の問題点を指摘しつつその論文の提案手法の特色を強調した文が含まれることが多く，本研究ではこのような文を「関連研究」のセグメントから抽出すべき重要文であると定めた．「関連研究」のセグメントからの要約文の抽出には，まず，関連研究に関する内容であると判定された段落に含まれる文を優先的に選択するようにした．続いて，関連研究の内容に関連する段落内の文，もしくはそれ以外の段落の文の重要度の優劣は，文中の単語の TF·IDF 値の和によって決定した．テストデータは，先に述べた「関連研究」のセグメントでの重要文の基準に適合する文を人手で選択し，作成した．重要文抽出の精度は，節のタイトルのパタンマッチで検出されたセグメント，手がかり句のパタンマッチで検出されたセグメント（段落）から重要文を抽出したとき，いずれも 20%程度であった．一方，再現率は，前者のセグメントが 32%，後者のセグメントが 22%で，前者の方が 10%程度高かった．最終的に，各セグメントから抽出された重要文を結合することで，包括的要約を自動生成することを提案した．

5.2 今後の課題

テストデータ 30 件を用いて行った評価実験の結果,「序論」のセグメントから抽出した重要文の精度，再現率，F 値はいずれも 30%程度であり，改善の余地がある．「序論」のセグメントからは，論文の目的，貢献について述べている文を抽出することが目的であるが,「序論」のセグメントから抽出すべき重要文の正解は，アブストラクトに含まれる文との類似度が高い文を自動的に選択している．したがって，論文の目的と貢献について述べている文か否か人手で判定して評価データを作成することが必要である．さらに，重要文かどうかを判定するモデルを学習するための素性として，文中に含まれる単語の n-gram(n=1,2,3) を用いるほか，論文の目的と貢献について述べている文であることを判定できる特徴の追加が必要とされる．

「関連研究」のセグメントから抽出した要約文の評価では，節のタイトルを手がかりとして検出されたセグメントからの重要文抽出と，関連研究の手がかり句によって検出されたセグメントとからの重要文抽出とでは，前者の再現率が高かった．これは，両者のセグメント検出の精度と関係があると考えられる．つまり，前者のセグメント検出の方が精度が高く，検出されたセグメントが正しいことが多いため，そのセグメントから重要文を抽出したときも正しい重要文を多く抽出できたと考えられる．「関連研究」からの重要文抽出の性能を向上させるためには,「関連研究」のセグメント検出の精度も高める必要がある．

「提案手法」「評価実験」のセグメントからの重要文抽出は未着手であるため，これを実現することは喫緊の課題である．3.3.3 項，3.3.4 項で述べた構想にしたがい，重要文，提案手法の概要を表わす図，実験結果を示した図や表などを抽出する手法を探究したい．

さらに，各セグメントから抽出した重要文を結合し，包括的要約を生成する手法も実装する必要がある．元の論文での出現順に重要文を結合したり，セグメント毎に重要文をまとめてから表形式で提示することで，ユーザが異なる観点から論文の要点を把握できるようにしたい．またユーザによって与えられた要約率を満たすように重要文を選択する手法は十分に検討する必要がある．このとき，各セグメント毎に要約率を満たすように選択するのか，あるいは「序論」のような重要と考えられるセグメントからより多くの重要文を選択するのかは，今後検討する必要がある．

以上で挙げた提案手法の改良や未実装のモジュールの実現に加えて，提案した手法で作成された包括的要約が実際のサーベイにどの程度役に立つのか，すなわち包括的要約の生成が複数の論文の内容を短時間で把握するのにどれだけ貢献するかを確認するための被験者実験も必要である．

参考文献

[1] Nitin Agarwal, Kiran Gvr, Ravi Shankar Reddy, and Carolyn Penstein Rosé. Towards multi-document summarization of scientific articles: Making interesting comparisons with scisumm. In *Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, WASDGML '11, pp. 8–15, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[2] Regina Barzilay and Kathleen R.McKeown. Sentence fusion for multidocument news summarization. *2005 Association for Computational Linguistics*, Vol. 31, No. 3, pp. 297–328, 2005.

[3] John M. Conroy and Dianne P. O'leary. Text summarization via hidden markov models. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 406–407, 2001.

[4] H. P. Edmundson. New methods in automatic extracting. *Journal of the ACM*, Vol. 16, No. 2, pp. 264–285, 1969.

[5] Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, and Masaaki Nagata. Single-document summarization as a tree knapsack problem. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1515–1520, 2013.

[6] Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 68–73, 1995.

[7] Chin-Yew Lin. Training a selection function for extraction. In *Proceedings of the Eighth International Conference on Information and Knowledge Management*, CIKM '99, pp. 55–62, New York, NY, USA, 1999. ACM.

[8] Chin-Yew Lin and Eduard Hovy. Identifying topics by position. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, ANLC '97, pp. 283–290, Stroudsburg, PA, USA, 1997. Association for Computational Linguistics.

[9] Annie Louis, Aravind Joshi, and Ani Nenkova. Discourse indicators for content selection in summarization. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '10, pp. 147–156, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[10] Ani Nenkova and Kathleen McKeown. Automatic summarization. *Foundations and Trends in Information Retrieval*, Vol. 5, No. 23, pp. 103–233, 2011.

[11] 奥村学, 難波英嗣. テキスト自動要約に関する研究動向. 自然言語処理, Vol. 6, No. 6, pp. 1–26, 1999.

[12] David Zajic, Bonnie J.Dorr, Jimmy Lin, and Richard Schwartz. Multi-candidate reduction: Sentence compression as a tool for document summarization tasks. *Information Processing & Management*, Vol. 43, No. 6, 2007.