氏　　名	関　洋平		
学位（専攻分野）	博士（情報学）		
学位記番号	総研大甲第859号		
学位授与の日付	平成１７年３月２４日		
学位授与の要件	複合科学研究科　情報学専攻 学位規則第６条第１項該当		
学位論文題目	文書ジャンルとテキスト構造に着目した自動要約		
論文審査員	主　査	教授	神門　典子
		教授	東倉　洋一
		助教授	影浦　峡
		教授	安達　淳（国立情報学研究所）
		教授	島津　明（北陸先端科学技術大学院大学）
		助教授	森　辰則（横浜国立大学）

論文内容の要旨

本論文では，文書ジャンルとテキスト構造を用いた自動要約の新たな手法を提案し，その有効性を検証した．文書ジャンルとは，日記や報告書のような文書の種類を意味する．また，テキスト構造は，テキスト全体に対する各文の果たしている役割や機能を構成要素として捉えてテキストの構成を説明する機能構造に着目した．このような構造は，その文書ジャンルのテキストに含まれる典型的な構成要素，テキストにおける構成要素の一般的な出現順序，構成要素間の関係などによって決まる．本研究では，この文書ジャンルとテキスト構造を要約作成のための元文書からの重要箇所抽出と要約文生成に用いた．

従来の自動要約研究では，主として文書中での内容語の出現頻度等を手がかりとして文書の話題に着目した要約作成を目指してきた．これらの研究は，作成された要約が内容のバランスと文章としての一貫性に欠け，また，ある話題について，たとえば事実を知りたいのか，意見を知りたいのかといった利用者が重視する情報のタイプの違いなどを扱うことはできず，利用者の情報要求における話題以外の側面に柔軟に応じることが難しいという問題があった．

それに対し，本論文は，文書の話題だけではなく，情報のタイプ（例えば，意見か，事実報告かなど）をも区別した要約作成という新しい課題を提起し，内容語の出現頻度とともに，文書ジャンルとテキスト構造に着目した新たな手法を提案した．そして，実験によって，提案手法が，利用者の重視する話題と情報のタイプ（例えば，意見を知りたいのか，事実を知りたいのかなど）に応じた要約作成，元文書からのバランスよい内容抽出，要約文の一貫性の向上において有効であることを示した．具体的には，以下の3点について研究を行なった．

A) 元文書からバランスよく内容を抽出した要約の作成

B) どういう種類の情報を知りたいか（意見を知りたいのか，事実を知りたいのか，知識を知りたいのか）を区別した要約作成

C) 書き手や読み手にとって自然な構成をもった要約文の生成

本論文を構成する研究テーマとして，以下の3つを取り上げた．

A. テキスト構造を用いた単一文書要約：解説記事の携帯電話向け要約を例として

研究 A と研究 B では，元文書における語の出現頻度とともに，文書ジャンルとテキスト構造を用いて要約に含める重要箇所を抽出する自動要約手法について取り組んだ．

ここでは，その予備的検討として，新聞の解説記事という文書ジャンルをとりあげ，特定の単一の文書ジャンルを対象とした単一文書要約作成器を作成した．テキスト構造の構成要素として，新聞記事に特徴的に現れる主記，解説，背景，見通し，意見という5つの文タイプを元文書中の文に自動付与し，それらを従来の重要箇所抽出の手法と組み合わせる要約作成手法を提案し，その有効性を調べた．新聞の解説記事のテキスト構造は，特定の話題と意見の繰り返しから構成されることから，本研究では，この構成を考慮し，主記と解説と意見とを組み合わせて要約を作成する戦略を採用した．

評価は，元記事の内容をたずねる質問の集合を用意し，要約だけを読んで正しく解答

できるかを調べた．その結果，提案手法であるテキスト構造を用いて作成した要約は，従来手法のリード法，および，語の頻度などを手がかりとして重要箇所を抽出する要約作成手法によって作成した要約に比べ，質問への解答率が，10要約の平均で，それぞれ5.8%，8.1%向上した．テキスト構造は要約作成に有望な傾向を確認し，次の B ではより複雑な複数文書要約への適用を試みた．

B. 文書ジャンルとテキスト構造を用いた複数文書要約：多様な文書ジャンルを対象とした利用者の観点に基づく複数文書要約

要約の利用者は，同じ話題に関する要約でも，あることがらや出来事の一連の経緯などの事実が知りたいのか，その出来事に関する識者の意見や書き手の意見を知りたいのかなど，利用者によって重視する情報のタイプが異なる．本研究では，この問題を解決するために，利用者が指定した要約の観点に応じて，複数文書要約を作成するシステムを実現した．

要約の観点にはさまざまな捉え方があるが，本研究では，利用者が重視する話題だけではなく，利用者が重視する情報のタイプも区別して，利用者の情報要求により適した要約を作成するために，文書ジャンルとテキスト構造を用いた．文書ジャンルは，詳細描写性，議論性，非個人性，事実性の4つの次元の値の組み合わせで表現するジャンル特性を設定した．テキスト構造の分析は，主記，解説，背景，著者意見，識者意見，見通しの6つの文タイプを使用した．また，提案手法の有用性を検討するために，利用者が重視する情報のタイプを指示して作成した人手作成参照要約を持つ要約実験用データセットViewSumm30を作成し，評価実験を行なった．

その結果，提案手法は，情報のタイプを考慮しないベースラインシステムよりカバレッジが向上し，その向上率は，事実報告型，意見重視型，知識重視型の要約について，30文書集合の平均で，それぞれ 5.4%，33.6%，24.6%であった．また，元文書中の事実，意見，知識を問う質問の集合を作成し，要約を読んだだけで解答できるか調べたところ，提案手法は，ベースラインシステムと比べて，解答率が有意に向上した．これにより，文書ジャンルとテキスト構造を用いることで，従来区別できなかった利用者の情報要求を区別した要約が作成できることがわかった．

C. 要約文生成におけるテキスト構造の利用：3段階モデルによるデータからの特定文書ジャンルの要約文生成

本研究では，従来の語の出現頻度や出現位置を手がかりとして重要箇所を抽出する要約が一貫性に欠けるという問題を解決するために，要約文作成においてテキスト構造を利用し，時間や場所などの手がかりにより情報を集約してテキストを生成する研究に取り組んだ．AとBの研究が，要約に含める重要箇所の抽出に，入力となる元文書のテキスト構造を利用していたのに対し，この研究では，出力する要約のテキスト構造に着目して情報をわかりやすくまとめて提示するテキスト生成技術を提案した

ここでは，数値データなどの抽象的な情報を簡潔にまとめて提示するために，テキスト構造を利用した．また，文書ジャンルの一例としては，天気予報を選択した．テキスト構造の表現にはXMLを用いた．

本システムにより自動生成した天気予報と実際の天気予報との一致について評価した結果，場所の表現について何らかのかたちで一致した天気予報が生成できることを確認した．また，テキスト構造を用いることで，大文字と小文字の区別や助動詞の活用などの生成処理を柔軟に実現し，読みやすい要約文生成を実現することができた．

AとCでは，単一の文書の要約を対象としており，文書にどういう種類の情報が現れるかをとらえるためにテキスト構造に着目した．それに対し，Bでは，複数の文書の要約を対象としているため，文書群にどういう種類の情報が現れるかという点について，各文書の文書ジャンルを区別している．ここでは，テキスト構造に加えて，文書ジャンルについても，被験者間の付与の一貫性について調査し，機械学習による自動付与を行なった．

本論文を構成する3つの研究では，それぞれ複数の被験者の付与実験またはコーパス文書の分析によってテキスト構造を構成する構成要素が利用者に共通に認識されることを確認し，要約作成への有効性について検証した．その結果，各研究について，テキスト構造を用いることの有効性を示すことができた．

論文の審査結果の要旨

関洋平君の学位申請論文は、文書ジャンルとテキスト構造を用いた自動要約の新たな手法を提案し、その有用性を検証したものである。自動要約は大量の電子的情報の活用支援技術として、近年、その重要性が増している。従来の自動要約研究では、主として文書中での内容語の出現頻度等を手がかりとして文書の話題に着目した要約作成を目指してきた。それに対し、本論文は、文書の話題だけではなく、情報のタイプ（例えば、意見か、事実報告かなど）をも区別した要約作成という新しい課題を提起し、内容語の出現頻度とともに、文書ジャンルとテキスト構造に着目した新たな手法を提案した。そして、実験によって、提案手法が、利用者の重視する話題と情報のタイプ（例えば、意見を知りたいのか、事実を知りたいのかなど）に応じた要約作成、元文書からのバランスよい内容抽出、要約文の一貫性の向上において有効であることを示した。

提出された論文は、全6章からなる。第1章では、本研究の動機、背景、目的を述べている。第2章では、関連する既往研究を整理し、それらを踏まえて、本研究の課題を再定義している。

第3章から第5章までは、提案手法を用いた自動要約の実験と評価結果である。第3章と第4章は、元文書における語の出現頻度とともに、文書ジャンルとテキスト構造を用いて要約に含める重要箇所を抽出する自動要約手法についてである。第3章は、その予備的検討として、新聞の解説記事を例としてとりあげ、特定の単一の文書ジャンルを対象とした単一文書要約を実装した。自動作成された要約と専門家が元文書から抽出した重要箇所との一致度、および、要約だけを読んで元文書の内容をたずねる質問群に正しく解答できるかという二つの面から評価を行ない、提案手法が、従来手法のリード法と重要箇所抽出法よりも有望であることを確認した。第4章では、より複雑な問題として、提案手法を複数の多様な文書ジャンルを対象とした複数文書要約に適用した。要約対象となる新聞記事群ごとに、話題と重視する情報のタイプとを指示して作成した人手作成参照要約と元文書の内容を問う質問群を持つ複数文書要約実験用データセットViewSumm30を作成し、これを用いて、提案手法と情報のタイプを考慮しないベースラインシステムとを比較した評価実験を行ない、その結果と考察を報告している。第5章では、出力する要約のテキスト構造に着目した要約生成技術を提案した。ここでは、大量の数値データを対象とし、生成する要約文書のジャンルの一例として天気予報を選択し、提案手法により自動生成した天気予報と実際の天気予報との一致について評価した結果を報告している。第6章で全体の結論を述べている

自動要約研究では、従来、主として要約対象となる元文書中に出現する内容語に着目して、文書の話題という観点から、要約に含める重要箇所を抽出するアプローチを採用してきた。これは文書の種類、長さ、言語、用途によらない汎用的な要約作成を目指すものである。それに対し、本研究は、従来手法に加えて、文書ジャンルとそのテキスト構造に着目することによって、バランスのよい一貫性の高い要約作成を指向するとともに、意見か事実報告かなどといった情報のタイプの違いに着目して、利用者の情報要求にそくした柔軟な要約生成を目指した点で高い独創性と新規性が認められる。本研究はこのような新しい要約研究の課題を提起し、その課題を解決するために語の出現頻度と

ともに文書ジャンルとテキスト構造にも着目した新しい要約作成のための元文書からの重要箇所の抽出と要約文生成との手法を提案し、実装し、実験を行ない、人間が作成した参照要約との対比による内的評価と要約の用途を指向した外的評価とによって有効性を検証している。また、本提案手法は、実際の情報検索の場で問題となっている漠然とした情報要求をもった利用者を、対話型で支援しながら情報要求を明確にし、検索結果をわかりやすく提示するための基礎技術として実用性の高い研究であり、今後の展開が期待される。本研究の成果の一部は、学術雑誌論文1篇、英文単行書収録論文2篇、査読付き国際会議論文2篇にまとめられている。そのほか、査読付き国際会議短報および国際会議併設ワークショップ論文3篇、国際会議併設ワークショップ招待講演、査読のない国際会議論文2篇、国内学会での口頭発表7篇があり、研究業績も十分である。予備審査において課題とされた諸点についても、本論文で十分に改善が図られていることが確認された。

以上から、博士論文審査委員会は、出願者の学位申請論文が博士（情報学）を授与するに値すると全員一致で判定した。