利用者の情報要求を考慮した観点に基づく複数文書要約とその評価

関 洋平[†,††,☆] 江口 浩二[††,†] 神門 典子[††,†]

本研究の目的は，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の実現である．要約の観点にはさまざまなとらえ方があるが，本稿では，ある文書集合中で利用者が関心を持ったトピック（話題）と利用者が重視する情報のタイプの 2 つに着目した．本システムは，文書集合が与えられると，その文書集合に含まれるトピックのリストを提示する．利用者がこのリストから興味のあるトピックを選び，重視する情報のタイプに応じて，事実報告型，意見重視型，知識重視型のなかから適当なものを選択すると，システムは，利用者が選択したトピックと情報のタイプを重視して当該文書集合の内容をまとめた要約を作成する．要約作成において重視するタイプの情報を抽出するための手がかりとして “文タイプ” と “文書ジャンル” を使用した．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，評価実験を行った．Closed な実験ではあるが，提案手法の，情報のタイプを考慮しないベースラインシステムに対するカバレッジの向上率は，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%だった．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．

Multi-document Viewpoint Summarization Based on Users' Information Needs and Its Evaluation

Yohei Seki,[†,††,☆] Koji Eguchi[††,†] and Noriko Kando[††,†]

The purpose of this study is to build a multi-document summarizer depending on user-specified summary viewpoints. Once a set of documents is provided to our system, a list of topics discussed in the given document set is presented, so that the user can select a topic or topics of interest as well as the information type to focus on in the summaries, such as facts, opinions, or knowledge, according to the user's information needs. We assume that sentence types and document genres are related to the types of information included in the source documents and they are useful to differentiate appropriate information types to focus on in the summaries. We evaluated the effectiveness of our proposal with the experimental dataset ViewSumm30. We found improvements of 5.4%, 33.6% and 24.6% for the coverage in fact-reporting, opinion-focused, and knowledge-focused summaries, respectively, compared with our baseline system. For extrinsic evaluation, we prepared sets of questions which can be answered only reading to ask facts, opinions, or knowledge described in the original documents for each of the 30 document sets. The correct answer rates are significantlly increased on the summaries produced by the proposed system than those by the baseline.

1. はじめに

我々は，利用者が指定した観点に基づき，複数文書からの要約を作成する研究を進めている．その一環として，利用者が選択したトピック（話題）に焦点を当てて要約を作成する複数文書要約器を実現した[1)]．本稿では，トピックに加えて，利用者が重視する情報のタイプを指定して要約を作成する複数文書要約器（Viewpoint Summarizer With Interactive clustering on Multiple documents：以降 v-SWIM とする）を提案する．

ある事柄やトピックについて情報を検索するとき，それぞれの利用者は，検索の目的や置かれている状況によって，その事柄やトピックのどのような側面に関心があるかが異なる[2)]．たとえば，“楽天のパ・リーグ参入” について調べるとき，“いつ，どこでどのようなことが起こったのか” という参入決定までの一連の事実を知りたいのか，利用者自身が参入の是非を判

† 総合研究大学院大学情報学専攻
Department of Informatics, The Graduate University for Advanced Studies
†† 国立情報学研究所
National Institute of Informatics
☆ 現在，豊橋技術科学大学
Presently with Toyohashi University of Technology

断する根拠として “楽天の経営状況” を示すデータなど確立した知識がほしいのか，あるいは，“仙台市民の評判” のような個人の主観を反映した陳述に関心があるのか，など，利用者の関心はさまざまである．そこで，本稿では，文書集合が与えられると，その内容を解析して文書集合中に含まれるトピックのリストを利用者に提示し，利用者が，(1) 興味のあるトピックと，(2) 重視する情報のタイプ（事実，意見，知識のいずれか）とを指定することで，利用者の興味にあった要約を提供する複数文書要約器を提案する．また，提案手法の有用性を検証するために，重視する情報のタイプを区別した人手作成参照要約（以下，参照要約）を持つ要約実験用データセット ViewSumm30 を構築し，評価実験を行った．

以下，2 章は要約の観点と利用者が重視する情報のタイプについて本稿の提案の位置づけを示す．3 章は実験方法である．4 章は要約作成において，元文書中の重視する情報のタイプを識別する手がかりについて説明する．5 章は実験結果と考察，6 章は結論である．

2. 提案：要約の観点を区別した複数文書要約

2.1 要約の観点：トピックと情報のタイプ

要約は，原文の内容を重要な主旨に絞って簡潔に表現する手続きである．Mani[3]（p.169）は，複数文書要約は，関連した文書集合に対して，利用者（またはアプリケーション）の要求に応じて，最も重要な内容を簡潔に提示する手続きと定義している．

利用者の情報要求によって “要約の観点” は異なる．“観点” とは，“観察，考察するときの立場や目の付けどころ” である[4]．本研究では，要約の観点として，トピックと，利用者が重視する情報のタイプに着目し，重視する情報のタイプとして，事実，意見，知識の 3 つを設定した．要約の観点のうちトピックについては，システムが与えられた文書集合の内容を解析して提示するトピックのリストから，利用者が選択した 1 つ以上のトピックに焦点を当てた複数文書要約を作成する機能を実現した[1]．本稿では，これに加えて，重視する情報のタイプを指定することで，利用者が選択したトピックと情報のタイプという観点を区別した複数文書要約を提案する．たとえば “楽天のパ・リーグ参入” というような，利用者が持つ比較的漠然としたクエリ（検索質問）に対する検索結果集合に相当する文書集合が与えられると，本稿の提案では，そこに含まれるより詳細なトピックや情報のタイプを利用者に提示して選択を求めることにより，利用者は最初に投入したクエリのどの側面に関心があるのかを明らかにし，利用者の潜在的な情報要求や観点により適合する要約を提供する．

2.2 先行研究に対する本研究の位置づけ

要約に対する利用者の情報要求☆は，クエリ[6] や質問[7] に焦点を当てるものが多い．クエリと質問との違いは，クエリの場合は，検索要求として指定したクエリに適合する文を要約に含むのに対し，質問は，その解答を要約に含むという点で異なる．本稿では，要約に対する情報要求としてのクエリと質問については区別せずに取り扱う☆☆．

- クエリが示す観点に応じた要約（Query-biased Summarization）
  Tombros ら[6] は，利用者が検索要求として表現したクエリに焦点を当てて，要約を作成し分けることを提案した．この手法では，トピックのように検索対象中に内容語として表現される要素に焦点を当てて要約を作成することができるが，事実，意見，知識などの情報のタイプという要素に焦点を当てて要約を作成し分けることはできない．
  これに対し本稿では，要約の観点として利用者が文書集合中のトピックと重視する情報のタイプとを指定する方法を提案する．重視する情報のタイプに応じた要約を作成するには，たとえば，“意見” という文字をクエリとしたり，“意見” という文字列の出現する箇所を取り出すのではなく，意見を述べている箇所を識別したりする手法（4 章）が必要となる．
- 観点記述に焦点を当てた要約
  Document Understanding Conference（DUC）2003[8] では，“観点記述に焦点を当てた要約” がタスクとして取り上げられた．このタスクでは，要約作成者の記述した要約の観点が，要約作成の手がかりとして参加者に与えられた．観点記述は，たとえば，“アメリカにおけるホームレスの程度，原因，帰結は何か”（d117：ホームレス），“国勢調査に対するサンプリングを使用した賛成と反対についての議論”（d115：国勢調査）など，文書集合中のトピックだけではなく，原因，帰結や議論などの書き手の主観的な判断や意見などを含む情報のタイプにも焦点が当てられていた．
  これに対し本稿では，トピックごとに個別の観点を記述するのではなく，重視する情報のタイプを観点の要素とし，情報要求を識別する．

☆ 関連研究は奥村ら[5]（4 章）や Mani[3]（3.3 節）も詳しい．
☆☆ 本研究では，後述する外的評価に使う質問はクエリとして入力することは想定しない．

2.3 利用者が重視する情報のタイプ

本稿では，“要約の観点” のうち，特に “トピック” と，“事実報告”，“意見”，“知識” などのどういうタイプの情報を重視するかという点に着目する．それぞれの情報のタイプを重視した要約は以下のように定義する．

(1) 事実報告型要約：発生した事件など過去に起こった事実を重視し，トピックに関連した事実報道が得られる要約．
(2) 意見重視型要約：著者の主張や識者の意見などの情報を重視し，トピックに関わるアドバイスや見通しや他者の評価などの情報を得られる要約．
(3) 知識重視型要約：事件よりも，すでにある定義的または事典的な知識を得られる要約．

“事実報告型要約” は，重要な事実に焦点を当てて作成する要約である．従来の要約研究では，報道記事などを対象として，内容語を手がかりとすることで，重要なトピックについての事実報告型の要約を作成し，どれだけの事実がカバーされているかを評価していた．以下では，それ以外の情報のタイプに焦点を当てた要約について，関連研究を示す．

- 主観的な記述を重視した要約
  Cardie ら[9] は，質問に対して複数の見方を区別した解答を提供するために，意見指向型要約（opinion-oriented summary）を用いることを提案している．Beineke ら[10] や Pang ら[11] は，情緒を重視した要約（Sentiment Summarization）を提案している．入力として質問は想定していない．
- 定義情報を抽出する要約
  Xu ら[12] は，定義をたずねる質問に対して，情報抽出の手法を用いて要約を作成している．Blair-Goldensohn ら[13] は，定義を表す述語を抽出して要約を作成している．藤井ら[14] は，事典的 Web 検索サイトにおける情報の集約に，複数文書要約技術を応用して説明情報を提供している．

これらの研究に対して，本稿の提案では，情報要求が明確でない利用者が，漠然としたクエリで検索したある大きなトピックに関する文書集合が与えられると，システムがその文書集合の内容を解析してトピックのリストを提示する．利用者は，システムが提示する候補からトピックを選択し，情報のタイプを指定することで，利用者の情報要求に関連した質問に解答可能な要約を作成して利用者を支援することを目指す．

本稿では，提案手法の有用性を評価するために，重視する情報のタイプに対応する質問に答えられるか調査した．Pomerantz[15]（pp.69–70）は，解答の粒度と機能によって以下の質問のタイプを挙げている：

- 短い解答を要求する質問のタイプ
  検証（真偽や事件の発生を問う質問），選択，概念補完（誰，何，いつ，どこを問う質問），特徴を問う質問，量を問う質問
- 長い解答を要求する質問のタイプ
  定義，事例，比較，解釈，原因，帰結，目標，手続き，可能化，予測，判断，表明，要求

これらのタイプの定義や用例を詳細に検討すると，解答としての重視する情報のタイプと以下のように対応づけることができる：

- 検証，概念補完，量を問う質問 → 事実報告
- 解釈，予測，判断，原因 → 意見
- 特徴を問う質問，定義，事例，手続き → 知識

これらの質問による要約の評価結果は，5.3 節で示す．

Pomerantz の質問のタイプのうち，“要求” は，解答者の行為を要求することから対象としない．そのほかの “選択，比較，帰結，目標，可能化，(わからないことの）表明” は，トピックに依存した質問の内容に応じて，特定の事実や意見や知識を要求する．このため，本稿でのトピック横断的な要約のための情報のタイプでは，これらの区別はカバーしない．

2.4 利用者が重視する情報のタイプを識別する手がかり

利用者が重視する情報のタイプに応じて複数文書要約を作成するために，本稿では文タイプと文書ジャンルを手がかりとして用いた．文タイプは，“背景” や “意見” などといった文の機能を分類したものであり，単一文書要約において情報要求の区別に有用であった[16)~18)]．文書ジャンルは，“日記” や “報告書” のような文書の種類を意味している．Bazerman[19] は，文書ジャンルを，“社会活動におけるコミュニケーションの役割を反映した形式” と定義し，Finn ら[20] は，“類似したスタイルで書かれた文書のグループで，トピックとは直交する（orthogonal）” 属性であると定義した．

本研究では，複数文書要約の元文書について，その文書の種類（社説，報道記事，特集記事など）に応じて，中心となる記述が事実，意見，知識といった点で異なることに着目し，まず文書ジャンルを情報のタイプを区別する手がかりとした．また，著者の主張が中心となる社説は，意見をサポートする事実のような要素が含まれている．これらのより細かい単位での情報のタイプを区別するために，形式的に取り扱いやすい

文を単位とする文タイプを手がかりとした．この着想の背景には，抄録作成の専門家は，文書の種類に応じたテキストの構成を利用して要約を作成するという報告[21]がある．

既存の新聞記事のジャンル分類として，IPTC（International Press Telecommunications Council）のメタデータがあるが，これは記事の種別，主題，ニュースソースなど多様な側面の分類を一次元に配列してあり，分類基準が複雑で人手付与も自動付与も困難である．そこで，本稿では文書ジャンルを，Biber の多次元分析[22]に準拠した 4 つの次元の値の組合せで規定した．各次元は，文書ジャンルを区別する属性を表す．本稿では各次元の値を組み合わせたものをジャンル特性と呼ぶ（4.2 節で詳述）．

3. 実験方法

本実験の目的は，2 つある．1 つは，利用者が重視する情報のタイプを考慮した複数文書要約について，文タイプと文書ジャンルの手がかりとしての有効性を検討することである．もう 1 つは，元文書の内容を問う質問を用いて，作成された要約の有用性を評価することである．前者については，評価手法としてカバレッジ（coverage）と精度（precision）を用いて人間が作成した要約への類似性について調査し，後者については，質問集合に対する解答率を計算した．これらの評価尺度については，3.3 節で報告する．

3.1 ベースラインシステム

ベースラインシステム☆は，利用者が指定したトピックに応じて要約を作成するが，利用者が重視する情報のタイプを考慮しない．完全リンク法，グループ平均法などの複数の文書クラスタリング手法の要約作成における有効性を比較した結果[1]，Ward 法を用いた段落単位のクラスタリングを採用した．アルゴリズムを下記に示す．

(1) 段落クラスタリング
  (a) 元文書を段落単位で分割し，それぞれの段落に対する単語の出現頻度を計算する．
  (b) すべての段落について，単語の頻度を利用した特徴素ベクトル間の距離に基づき，クラスタリングを行う．クラスタの数は，抽出文の数に基づいて決定する．

(2) 文抽出
  (a) 各クラスタの特徴素ベクトルを，単語の出現頻度と出現クラスタの頻度の逆数を掛け合わせた値☆☆で重み付けする．
  (b) 要約に焦点を当てる（利用者が指定した）トピックの有無で条件分岐を行う．
    (i) トピックがある場合
      質問中の単語と，各クラスタの特徴素ベクトルとの類似度から，クラスタの順位をつける．
    (ii) トピックがない場合
      全文書の単語の出現と，各クラスタの特徴素ベクトルとの類似度から，クラスタの順位をつける．
  (c) 各クラスタ中の文を，クラスタに含まれる元文書の見出し中の単語，クラスタにおける単語の出現頻度などを利用して重み付けする．
  (d) クラスタの順位に基づき，順位の高いクラスタから順に重みの大きい 1〜2 文を抽出する．抽出文数または要約文字数に達した段階で停止する．

3.2 提案手法の実現

ベースラインシステムは，利用者が指定したトピックに応じて要約を作成する機能を持つ．提案手法は，これに加えて，利用者が重視する情報のタイプも考慮した要約を作成することができる．その手がかりとして，ここでは，文タイプとジャンル特性（4 章で詳述）を用いた．

提案手法では，元文書中の文に文タイプを，文書にジャンル特性を自動付与し，文タイプ付き文とジャンル特性を持つ記事中の文の重み（前節のベースラインシステムのアルゴリズムの (2)(c) で計算したもの）にバイアスを設定することで，重視する情報のタイプを考慮した要約を作成した．

また，5.2 節で示すように，手がかりに対するバイアスは，参照要約に対する類似性の評価尺度であるカバレッジを使用し，重視する情報のタイプごとに最適なバイアスを実験により経験的に計算した．

3.3 重視する情報のタイプを考慮した要約作成重視する情報のタイプを区別した実験用データセット

本稿で提案する観点に基づく複数文書要約の有効性を評価するために，実験用データセット ViewSumm30 を構築した．これは，30 の課題☆☆☆があり，それぞれ，

☆ NTCIR-4 Text Summarization Challenge（TSC）3[23)〜25)] における評価については Seki ら[1] を参照．

☆☆ 単語の出現頻度 * log（全クラスタ数/出現クラスタ数）

☆☆☆ なお，DUC や TSC では “topic” といわれているが，ここでは観点の一要素として指定しているトピックと区別するために，課題内容と呼ぶ．課題内容と ID の組合せを課題と呼ぶ．

表 1 重視する情報のタイプを区別した実験用データセット ViewSumm30
Table 1 Experimental dataset ViewSumm30 with information types.

課題		文書集合		参照要約の長さ		
ID	課題内容	文字合計	記事数	事実	意見	知識
S010	欧州通貨統合	20530	10	786	795	790
S020	年金支払い抑制	21704	10	787	789	788
S030	粉飾決算	21207	9	798	796	791
S040	イトマン事件	20647	10	791	798	784
S050	ペイオフ解禁	19251	11	783	799	777
S060	次世代デジタル携帯	20353	11	798	797	782
S070	ガイドライン関連法	20687	9	776	785	799
S080	コソボ	20583	11	800	767	772
S090	戦略兵器削減	15499	8	764	765	796
S100	脳死判定	21052	7	795	776	785
S110	少年審判	20967	11	764	788	800
S120	情報公開法	16953	8	775	760	759
S130	ドナーカード	15902	10	764	797	738
S140	確定拠出型年金	19131	12	761	769	760
S150	遺伝子組み換え食品	20225	12	799	763	784
S160	組織犯罪対策法	21425	8	768	771	789
S170	臨界事故	16935	7	794	763	797
S180	金融ビッグバン	19411	8	762	795	795
S190	プルサーマル	19092	9	760	796	797
S200	戦域ミサイル防衛	17323	8	770	777	771
S210	中国国有企業	13529	6	763	766	707
S220	北アイルランド紛争	14241	10	790	780	770
S230	ロシア経済金融危機	15861	7	782	793	797
S240	テポドン	20130	8	797	795	783
S250	国際人権規約	20952	7	792	777	785
S260	大統領弾劾裁判	19170	8	791	790	783
S270	太陽政策	16942	7	796	799	787
S280	環境ホルモン	18368	10	790	787	792
S290	国際宇宙ステーション	15121	8	791	797	782
S300	世界遺産条約	16812	7	794	785	797
	最大	21704	12	800	799	800
	最小	13529	6	760	760	707
	平均	18666.8	8.9	782.7	783.8	781.2
	標準偏差	2323.4	1.6	13.6	12.9	19.3

課題内容に関連する新聞記事の集合がある．文書集合は，読売新聞と毎日新聞の合計で 10 記事程度から構成されており，各課題について，日本語の参照要約と内容を問う質問の集合を作成した．ViewSumm30 には，以下の特徴がある．

(1) 利用者が重視する 3 つの情報のタイプ（事実報告，意見，知識）を区別した．
(2) 元文書は，報道記事，社説，インタビュー記事など，さまざまな種類の記事を含む．

ViewSumm30 の課題内容と文書集合を**表 1** に示す．評価に使用する参照要約は，それぞれの課題について，同じ 1 人の作成者が，事実報告型，意見指向型，知識重視型の 3 つの要約を作成した．作成者は，専門誌や教科書の編集者 3 名である．それぞれが，15 課題，7 課題，8 課題の要約作成を担当した．

要約作成者への指示は，参照要約の長さはすべての課題について 800 文字以下とした．上限を 800 文字としたのは，圧縮率を考えるとこれ以下の長さにすることはむずかしく，要約を読む利用者としては，800 文字以上の長さは適当でないと判断したことによる．要約作成者には，該当するタイプの情報がない場合には，800 文字を大きく下回ってもよいと指示を与えたが，結果として，表 1 に示すように，どの課題と重視する情報のタイプについても 800 文字近くの参照要約が得られた．いくつかの課題は，要約作成者が，重視する情報のタイプ以外の要約の揺れを制限するために，要約を作成する際に着目した（サブ）トピックを指定している．このトピックは，3.1 節のアルゴリズムに基づき，ベースラインについても提案手法についても要約作成に利用しており，提案手法のベースラインに対する差には寄与していない．

評価手法と尺度

内的評価として，参照要約を基準として，本システムが抽出した文について，カバレッジと精度を計算した．Hirao ら[25] は，“精度” を，従来の解釈に従い，システム作成要約中の文の集合が，参照要約に対応する（元文書中の）文の集合に含まれている割合として，“カバレッジ” を，システム作成要約中の文の集合が，参照要約にどれだけ近いかを冗長性排除の性能を考慮して評価する尺度として定義した．本稿では，この尺度を用いて，それぞれの重視する情報のタイプについて有効な文タイプとジャンル特性を調査した．

次に，外的評価として，システムが作成した要約だけを読んで，元文書の内容を問う質問に答えられるか実験した．質問を利用した要約の外的評価は Hirao ら[26] などで提案されているが，本稿では，元文書中の事実，意見，知識を問う 3 種類の質問集合を用いた．質問は，参照要約を作成したうちの 2 名が，事実，意見，知識を問うものを，それぞれ各課題につき 3〜9 件作成した．作成者は，それぞれ参照要約を作成していない課題について，質問と解答を作成し，質問の優先順位をつけた．質問数は，30 課題の合計で，事実が 182 件，意見が 179 件，知識が 174 件となった．評価は，著者が解答を参照しながら行った．

作成した質問の集合については，以下の 2 種類の手順で解答率（Answer Rates）の評価を行った．

(1) 質問間の優先順位を考慮しない解答率
各質問に付与された優先順位を考慮せず，以下の式の値を，30 の課題について平均をとった．

$$\text{Answer Rates} = \frac{\#\text{ of Answers}}{\#\text{ of Questions}} \quad (1)$$

# of Answers：要約だけを読んで正しく答えることができる質問数
# of Questions：すべての質問数

(2) 質問間の優先順位を考慮する解答率
各質問に付与された優先順位の逆数（Reciprocal Rank）を各質問の得点として，以下の式の値を，30 の課題について平均をとった．

$$\text{Answer Rates} = \frac{\text{Scores of Answers}}{\text{Scores of Questions}} \quad (2)$$

Scores of Answers：解答の得点
Scores of Questions：質問すべての得点

どちらの解答率も，設定された質問すべてについて，完全に解答すると，そのスコアは 1 となる．

4. 情報要求を識別する手がかり

本章では，文タイプと文書ジャンルの，定義，人手付与の一致度，自動付与について説明する．

4.1 文タイプ

定　義

文タイプ[16)~18)]は，文を単位とした情報を識別する手がかりとして，使用されている．神門[27)]は，新聞記事に対する 5 つの文タイプとして，“主記”，“解説”，“背景”，“意見”，“見通し” を定義している．複数被験者間の付与実験では，一致率が各々89.1%，88.2%，96.4%，100%，80%と高い[27)]．ここでは，神門[27)]を拡張し，6 つの文タイプを使用した．本研究では，“著者の意見” と “識者の意見” は，付与の性質が異なると判断した．

(1) “主記”（M）：文書中で中心となる内容．
(2) “解説”（E）：主記を詳述．
(3) “背景”（B）：トピックに関連した歴史または背景を記述．
(4) “著者の意見”（O1）：記事の著者の意見．
(5) “見通し”（P）：将来起こりそうな出来事を表現．
(6) “識者の意見”（O2）：専門家などの第三者の意見として報道されたもの．

人手付与と付与者間の一貫性

まず，1994 年の日本経済新聞の記事 352 件（5,201 文）のすべての文について，神門[27)]の 5 つの文タイプを排他的に付与したデータを調査した．この付与は 2 名の被験者が独立に付与し，双方で協議を行い一致した結果を最終結果としていた．

しかし，“解説”，“背景” の中には，“見通し”，“意見” のタイプを兼ねている文があることから，“著者の意見”，“識者の意見”，“見通し” タイプの文について，他の文タイプとの排他性を考慮せずに新たに 2 名の被験者☆が，付与を行った．2 名の被験者間の付与の一致率を表すカッパ係数[28)]は，著者意見について 0.935，識者意見について 0.888，見通しについて 0.901 となり，非排他的付与について高い一致が見られた．

さらに，他の文タイプについても非排他的付与の観点から見直しを行った結果，352 件の記事に対する文タイプの付与数は，主記，解説，背景の付与と合わせて以下のようになった：

主記（922 文），解説（2,521 文），背景（1,447 文），著者意見（468 文），識者意見（856 文），見通し（593 文）．

自動付与

文タイプの自動付与は，機械学習である Support Vector Machines（SVM）[29)]を利用して実現した．SVM は過学習に対して頑健であり，多くの特徴素を取り扱える利点がある[29)]．本稿の自動付与で利用した特徴素を以下に示す．(8)，(9) については，Seki ら[16)]で用いた付与データを利用して特徴素を選択した．

(1) 文の文書内位置と段落内位置．
(2) 文を含む段落の文書内位置．
(3) 文の長さ．
(4) 文が含む，記事の見出しに現れる単語の数．
(5) 文が含む，TF-IDF☆☆の値が高い単語の数．
(6) 助動詞を手がかりとした態，時制，モダリティ．
(7) 南瓜[30)]を用いた 8 種類の固有表現の頻度．
(8) 分類語彙表[31)]を用いた 20–40 種類の用言，主語の基本意味素．
(9) 背景，著者意見，識者意見，見通しタイプのそれぞれに関連した 30–40 種類の句．
(10) 前後の文の文タイプ情報．

評価は，SVM を用いた自動分類研究で一般的に用いられている精度（Precision），再現率（Recall），正確さ（Accuracy）を用いた[29)]．文タイプ付与の評価手続きには，352 の記事に対する，k-fold cross-validation（$k = 4$）を用いた[29),32)]．結果を**表 2** に示す．精度と再現率のマクロ平均は，主記（精度 0.804，再現率 0.890），解説（精度 0.903，再現率 0.756），背景（精度 0.714，再現率 0.763），著者意見（精度 0.635，再現率 0.502），識者意見（精度 0.666，再現率 0.649），見通し（精度 0.697，再現率 0.483）となった．

この付与データを学習データとして，実験用データセット ViewSumm30 中の 11,931 文に対する自動付

☆ 最初の被験者とは異なる 2 名．
☆☆ TF-IDF は，TF は記事中の単語の出現頻度，IDF は新聞記事の集合を単位として計算した．

表 2 文タイプ自動付与の 4-fold cross validatoin を用いた正確さ，精度，再現率
Table 2 Accuracy, precision, and recall for sentence-type.

文タイプ	主記（M）			解説（E）			背景（B）		
	正確さ	精度	再現率	正確さ	精度	再現率	正確さ	精度	再現率
グループ A	0.954	0.809	0.916	0.815	0.918	0.664	0.846	0.740	0.816
グループ B	0.972	0.877	1.000	0.878	0.895	0.857	0.863	0.701	0.720
グループ C	0.869	0.646	0.656	0.798	0.910	0.657	0.845	0.707	0.694
グループ D	0.974	0.884	0.986	0.876	0.890	0.846	0.862	0.711	0.822
マクロ平均	0.942	0.804	0.890	0.842	0.903	0.756	0.854	0.714	0.763
マイクロ平均	0.929	0.777	0.843	0.830	0.904	0.726	0.851	0.719	0.766
ViewSumm30	0.967	0.671	0.813	0.733	0.603	0.673	0.824	0.700	0.833

文タイプ	著者意見（O1）			識者意見（O2）			見通し（P）		
	正確さ	精度	再現率	正確さ	精度	再現率	正確さ	精度	再現率
グループ A	0.916	0.739	0.527	0.893	0.661	0.705	0.918	0.871	0.449
グループ B	0.959	0.600	0.500	0.912	0.740	0.626	0.924	0.755	0.514
グループ C	0.897	0.486	0.368	0.843	0.594	0.489	0.873	0.451	0.446
グループ D	0.970	0.714	0.614	0.914	0.670	0.774	0.940	0.711	0.524
マクロ平均	0.936	0.635	0.502	0.891	0.666	0.649	0.914	0.697	0.483
マイクロ平均	0.927	0.621	0.470	0.882	0.647	0.620	0.907	0.626	0.469
ViewSumm30	0.873	0.761	0.531	0.895	0.844	0.759	0.968	0.669	0.272

与を行った．ViewSumm30 中の文の集合に対する正確さ，精度，再現率を表 2 の一番下の行に示す．

4.2 文書ジャンル

定　　義

本稿では，Biber らの提案[22] にヒントを得て，文書ジャンルを複数の次元の値の組合せとして定義した．2.3 節で述べたように，これをジャンル特性と呼ぶ．本稿で，ジャンルを複数の次元の値の組合せとして定義することの利点は以下のとおりである．

- それぞれの次元の効果を明確に検証できる．
- 付与の規則が単純化できる．
- 新しいクラス（ジャンル）を既存の次元の組合せで表現することができる．

本稿では，4 つの次元を下記のように設定した．これらは，本稿中で G1，G2，G3，G4 と参照する．

(1) **状況即時性**（+）対**詳細描写性**（−）（状況即時性，G1）
見た状況を即時的にありのまま描写している記事は，状況即時性を持つと判定する．後からふりかえった考察や分析など詳細な解説に記事が基づいている記事は，詳細描写性とする．

(2) **議論あり**（+）対**議論なし**（−）（議論性，G2）
ある主張の根拠となる事実や引用が含まれており，説得や予測などの目標に向けて論旨の展開がなされている記事は，議論性があると判定する．

(3) **非個人スタイル**（+）対**個人スタイル**（−）（非個人性，G3）
主語の不在，受動態などを手がかりとして，第 3 者から見た見地で書かれている記事は，非個人スタイルと判定する．著者が自分の意見を述べている場合は，個人スタイルとする．

(4) **事実性**（+）対**意見性**（−）（事実性，G4）
記事の文章構成を考慮したうえで，最も主張したい内容が事実か意見かに基づいて判定する．

表 3 カッパ係数：被験者間の付与の一致度
Table 3 Kappa coefficient among annotators.

ジャンル特性の次元	被験者の組			
	(a1,a2)	(a1,a3)	(a2,a3)	平均
状況即時性 (G1)	0.618	0.595	0.665	0.626
議論性 (G2)	0.410	0.536	0.678	0.541
非個人性 (G3)	0.459	0.506	0.604	0.523
事実性 (G4)	0.604	0.566	0.657	0.609

a1〜a3：各被験者による付与．

人手付与と付与者間の一貫性

ジャンル特性の人手付与の一致度を評価した．まず，3 名の被験者が，1998〜1999 年の毎日新聞と読売新聞の記事 208 件（ViewSumm30 の記事とは異なる）に対し，ジャンル特性の 4 つの次元について 0〜3 の間の整数の値を付与した．0 は，各次元の定義の（−）の特性（詳細描写性など），3 は，(+）の特性（状況即時性など）と被験者が判断した場合に付与した．明確に判断できない場合には，1 または 2 を付与した．

次に，3 名の被験者による付与の一致について，カッパ係数を調べた．カッパ係数は，3，2 を（+）の特性，1，0 を（−）の特性として計算した．結果を**表 3** に示す．カッパ係数の平均は，0.5 から 0.7 の間の値となった．Landis ら[33] の評価の基準によると，0.4–0.6 の値は “中程度（moderate）” であり，0.6–0.8 の値は “十分（substantial）” である．よって，人間の被験者が比較的一致して付与できることが確認できた．

自 動 付 与

ジャンル特性の自動付与についても，文タイプの自

表 4 ジャンル特性自動付与の 4-fold cross validatoin を用いた正確さ，精度，再現率
Table 4 Accuracy, precision, and recall of genre classification.

ジャンル特性	状況即時性（G1）			議論性（G2）			非個人性（G3）			事実性（G4）		
	正確さ	精度	再現率	正確さ	精度	再現率	正確さ	精度	再現率	正確さ	精度	再現率
グループ A	0.827	0.886	0.861	0.846	0.700	0.583	0.885	0.933	0.933	0.904	0.953	0.953
グループ B	0.788	0.852	0.767	0.885	0.727	0.727	0.750	0.826	0.884	0.923	0.953	0.953
グループ C	0.846	0.958	0.767	0.923	0.875	0.700	0.885	0.976	0.891	0.904	1.000	0.894
グループ D	0.750	0.722	0.897	0.904	0.500	0.800	0.942	0.957	0.978	0.923	0.957	0.957
マクロ平均	0.803	0.855	0.823	0.889	0.701	0.703	0.865	0.923	0.922	0.913	0.966	0.939
マイクロ平均	0.803	0.844	0.824	0.889	0.703	0.684	0.865	0.922	0.922	0.913	0.966	0.939
ViewSumm30	0.831	0.345	0.278	0.682	0.661	0.832	0.794	0.841	0.752	0.693	0.674	0.715

表 5 3 つの要約に対するカバレッジと精度の向上率
Table 5 Coverage and precision improvement effect for three type summaries.

	事実報告型			意見重視型			知識重視型		
	カバレッジ	精度	カバレッジ向上率	カバレッジ	精度	カバレッジ向上率	カバレッジ	精度	カバレッジ向上率
v-SWIM（理論的上限値）	0.210*	0.197	14.1 (%)	0.161**	0.144	46.4 (%)	0.197**	0.177	42.8 (%)
v-SWIM（自動付与）	0.194	0.182	5.4 (%)	0.147*	0.129	33.6 (%)	0.172*	0.161	24.6 (%)
ベースライン	0.184	0.171	-	0.110	0.099	-	0.138	0.140	-

動付与と同様に Support Vector Machines（SVM）を利用した．ここでは以下の特徴素を使用した．

(1) 記事特徴素：署名，記事の掲載された紙面，写真，図，ニュースソース（5 要素）．
(2) 統計的特徴素：文字数，異なり語数/総語数，文の数，意見文の数，見通し文の数，背景文の数，接続詞の数，引用括弧の数，平均文長（9 要素）．
(3) 機能語：“たい”，“らしい”，“感じられる” など意見，見通し，背景を表現する句（60 要素）．
(4) 記号：“，”，“．”，“「”，“」” などのパンクチュエーション（93 要素）．
(5) 固有表現：南瓜[30] を用いた固有表現の頻度（8 要素）．
(6) 用言，主語の意味素：分類語彙表[31] を用いた用言，主語☆の意味素の頻度（約 20 要素）．

ジャンル分類には，機能的語句や句読点記号などのパンクチュエーション[34] は有効であるといわれており，統計的特徴素[35] もよく利用される．ここでは，これらに加えて，記事特徴素を定義した．(3) は，4.1 節の，文タイプ自動付与の (9) の特徴素から選択した．(6) については，文タイプ自動付与の (8) の特徴素のうち，記事での出現頻度が高いものを利用した．

ジャンル特性自動付与の評価手続きは，208 記事に対する，k-fold cross validation（$k = 4$）を用いた．結果を**表 4** に示す．評価は，訓練データは，被験者 3 名が付与した値である 0〜3 について，0，1 を負（−）の特性として，2，3 を正（+）の特性として各次元の値を決定した．精度と再現率のマクロ平均は，G1（精度 0.855，再現率 0.823），G2（精度 0.701，再現率 0.703），G3（精度 0.923，再現率 0.922），G4（精度 0.966，再現率 0.939）となった．

この 208 件の新聞記事に対する付与データを学習データとし，ViewSumm30 の 267 記事についてジャンル特性の自動付与を行った．ViewSumm30 中の記事の集合に対する正確さ，精度，再現率を表 4 の一番下の行に示す．

5. 実験結果と考察

5.1 ベースラインシステムの評価

ベースラインシステムは，3.1 節で説明したシステムを使用した．ベースラインシステムの ViewSumm30 を用いた評価を**表 5** に示す．カバレッジは，事実，意見，知識という重視する情報のタイプについて，0.184，0.110，0.138 と低い．一方，このシステムを，NTCIR-4 TSC で使用された要約の実験用データセット☆☆で評価した際のカバレッジと精度は，Short が，0.376（0.419）と 0.471（0.591）で，Long が，0.429（0.433）と 0.535（0.587）であり☆☆☆，バグ修正前も修正後も参加チームで上位の結果であった．

要約などの情報アクセス技術の有効性の指標はデー

☆ 主語の意味素は G3 の自動付与の場合だけ使用．

☆☆ 報道記事が中心の元記事集合に対する複数文書要約の実験用データセットで，要約の元記事集合に対する圧縮率は約 5%と 10%，5%を Short，10%を Long と呼ぶ．

☆☆☆ 2004 年 10 月に明らかになった評価プログラムのバグを修正した後の値，修正前の値は，Hirao ら[25] の SOUKEN(a)(b) を参照．

表 6 ViewSumm30 と NTCIR-4 TSC の比較
Table 6 Comparing ViewSumm30 with NTCIR-4 TSC.

	ViewSumm30			NTCIR-4 TSC	
	事実	意見	知識	Short	Long
圧縮率 (%)（抽出文単位）	4.6	4.9	4.7	7.7	14.4
文書長（文数）（記事平均）	44.7			10.2	
抽出文の数（課題平均）	17.8	19.0	18.2	8.6	15.9
別の文でも正解となる解答の数（課題平均）	8.2	5.3	6.5	16.2	22.1
異なり語数（課題平均）	673.8			533.5	
名詞の異なり意味数（課題平均）	298.3			201.6	

タセットごとに異なるため，ViewSumm30 の値と NTCIR-4 TSC の値は比較できない．カバレッジの値が異なる原因は，いろいろ考えられる．

- 重視する情報のタイプを扱っていること．
- 抽出文の数に対する別の文でも正解となる解答の数の少なさ．
- 元文書集合の性質の差（冗長性）．
- ViewSumm30 の圧縮率が高いこと[36]．

ViewSumm30 と NTCIR-4 TSC のデータセットの性質の差を**表 6** に示す．カバレッジの差異の原因については，NTCIR-4 TSC のデータセットでは，タスクの焦点を複数の新聞社が発行する記事集合からの複数文書要約の冗長性排除の性能評価に当てていたため，正解の別解を多く設定している影響が考えられる．すなわち，ViewSumm30 では，最小の参照要約の文書長の 1 課題あたりの平均が約 17.8〜19.0 文であり，別解が 1 課題あたり約 5.3〜8.2 文程度である．これに対して，NTCIR-4 TSC のデータセットでは，Short は最小の参照要約の文書長の 1 課題あたりの平均が約 8.6 文（Long は約 16.2 文）であり，別解が 1 課題あたり約 16.2 文（Long は約 22.1 文）である．元文書集合中の内容の冗長性の差異は，異なり語数や名詞異なり意味数にも現れている．この問題については，今後の課題として，引き続き検討する．

5.2 提案手法の評価

3.2 節で述べた提案手法を用いて，3 つの重視する情報のタイプについて，ベースラインシステムよりもカバレッジの高い要約を作成した．情報のタイプごとの文タイプとジャンル特性を利用した重み付けを**表 7** に示す．表 7 の読み方は，1 行目が情報のタイプを表しており，2 行目以降は，それぞれの情報のタイプに対する（重み付けする特性，最適な重み，制約条件）の組合せを示している．制約条件は，重み付けする特性とあわせて成立する文タイプの条件を表している．制約条件がないときは “なし” と表記している．事実報告型要約の列の 5 行目は，“事実報告型要約を生成する際には，元文書中に現れる著者意見タイプで主記タイプではない文の重みを 0.1 倍する” と読む．各情報のタイプの最適な重みは，以下の手続きで決定した．

(1) 文タイプを利用した重み付け
- それぞれの情報のタイプに関連があると考えられる文タイプについて，0 から 5 までの間を 0.1 刻みで重み付けし，参照要約に対してカバレッジが最も大きくなるように値を調整した．
- 文タイプは非排他的付与を前提としており，他の文タイプの付与の有無を制約条件として組み合わせた．

(2) ジャンル特性を利用した重み付け
- それぞれの情報のタイプに関連があると考えられるジャンル特性について，0 から 5 までの間を 0.1 刻みで重み付けし，参照要約に対してカバレッジが最も大きくなるように値を調整した．
- ジャンル特性に特有の文タイプの付与の有無を制約条件として組み合わせた．

(3) 重み付けする文タイプとジャンル特性の数は，あわせて 6 つを上限とした．

このように重み付けした結果，向上したカバレッジと精度の値を表 5 に示す．まず，提案手法がどこまでよくなる可能性があるかを調べるために，人手で付与した文タイプとジャンル特性を用いて表 7 に従って文を重み付けして得られた要約のカバレッジと精度を，提案手法の理論的上限値とした．次に，それを自動化したらどこまで実現できるかを調べるために，自動付与した文タイプとジャンル特性を用いて表 7 に従って文を重み付けして得られた要約のカバレッジと精度を，提案手法の実際の評価とした．

表 5 中の *，** は，30 の課題について提案手法とベースラインを Wilcoxon の符号付順位検定で評価した結果，それぞれ有意水準 5%，1%で有意差があることを示す．カバレッジの向上率は，式 (3) で計算した：

$$\frac{\text{v-SWIM Coverage} - \text{Baseline Coverage}}{\text{Baseline Coverage}} * 100 \quad (3)$$

この実験は，Closed な実験であるが，提案手法がベースラインと比べて有望なことを示すことができた．

表 7 3 つの要約に対する文タイプとジャンル特性の重み付け
Table 7 Sentence weighting in the specific document genre for three type summaries.

事実報告型			意見重視型			知識重視型		
重み付けする特性	重み	制約条件	重み付けする特性	重み	制約条件	重み付けする特性	重み	制約条件
主記（M）	1.2	なし	解説（E）	0	識者意見ではない	解説（E）	4.5	主記ではなく 背景でもなく 著者意見でもない
解説（E）	1.2	なし	著者意見（O1）	1.7	なし	著者意見（O1）	0.2	解説ではない
著者意見（O1）	0.1	主記ではない	状況依存性（G1 正）	0	主記か背景	見通し（P）	0.2	なし
識者意見（O2）	0	主記ではない	議論性（G2 正）	1.7	主記か背景	状況依存性（G1 正）	0.4	なし
議論性（G2 正）	0.9	主記ではなく 解説でもない	事実性（G4 正）	0.3	主記	詳細描写性（G1 負）	2	解説か背景か 識者意見
非個人性（G3 正）	2	主記か解説か背景	意見性（G4 負）	1.5	識者意見で 背景ではない	意見性（G4 負）	0.1	なし

表 8 質問を用いた外的評価
Table 8 Extrinsic evaluation with questions for understanding original documents.

	事実		意見		知識	
	優先順位考慮	優先順位考慮せず	優先順位考慮	優先順位考慮せず	優先順位考慮	優先順位考慮せず
v-SWIM（理論的上限値）	0.495^{**}	0.407^{**}	0.328^{*}	0.279^{*}	0.554^{*}	0.468^{*}
v-SWIM（自動付与）	0.454^{*}	0.369^{*}	0.317^{*}	0.276^{**}	0.525	0.421
参照要約	0.640^{**}	0.569^{**}	0.427^{**}	0.382^{**}	0.622^{**}	0.469^{*}
ベースライン	0.350	0.285	0.227	0.157	0.436	0.351

提案手法の理論的上限値は，Wilcoxon の符号付順位検定を用いて，すべての重視する情報のタイプについて，ベースラインと比較して有意な向上があることがわかった．自動付与した文タイプとジャンル特性を用いた提案手法は，意見重視型要約と知識重視型要約について，ベースラインと比較して有意な向上があることがわかった．

事実報告型要約

事実報告型要約では，“主記”，“解説”，“著者意見”，“識者意見” の文タイプと “議論性”，“非個人性” のジャンル特性を組み合わせてバイアスをかけることで，カバレッジが 0.194，精度が 0.182 に向上した．カバレッジの向上率は，ベースラインと比較して 5.4%であった．

意見重視型要約

意見重視型要約では，“解説”，“著者意見” の文タイプと “状況依存性”，“議論性”，“事実性”，“意見性” のジャンル特性を組み合わせてバイアスをかけることで，カバレッジが 0.147，精度が 0.129 に向上した．カバレッジの向上率は，ベースラインと比較して 33.6%であった．

知識重視型要約

知識重視型要約では，“解説”，“著者意見”，“見通し” の文タイプと “状況依存性”，“詳細描写性”，“意見性” のジャンル特性を組み合わせてバイアスをかけることで，カバレッジが 0.172，精度が 0.161 に向上した．カバレッジの向上率は，ベースラインと比較して 24.6%であった．

さらに，作成した要約とは異なる情報のタイプの参照要約に対するカバレッジを計算した．

- 事実を重視して作成した要約は，意見重視型参照要約，知識重視型参照要約に対するカバレッジが，0.008，0.030 となった．
- 意見を重視して作成した要約は，事実報告型参照要約，知識重視型参照要約に対するカバレッジが，0.003，0.040 となった．
- 知識を重視して作成した要約は，事実報告型参照要約，意見重視型参照要約に対するカバレッジが，0.031，0.000 となった．

この結果から，それぞれの情報のタイプを区別して要約を作成していることが明らかになった．

5.3 質問応答を利用した元文書集合の内容の理解度をテストする外的評価

3.3 節で定義した 2 種類の解答率に基づき，事実，意見，知識に基づく要約について，参照要約，ベースライン，提案手法について評価した結果を**表 8** に示す．表 8 中の *，** は，30 の課題について提案手法とベースラインを Wilcoxon の符号付順位検定で評価した結果，それぞれ有意水準 5%，1%で有意差があることを示す．この結果により，本提案手法が，3 種類の情報要求を区別した質問集合に対して解答率が高い要約を，有意差を持って作成できることがわかった．提案手法により解答可能な質問の例としては，意見を問う質問である「少年審判のあり方を見直そうという動きに対して，どんな意見があったか」(S110)，「ロ

表 9 カバレッジが向上した課題数
Table 9 Number of topics improved for coverage.

	事実報告型		意見重視型		知識重視型	
	自動付与	理論的上限値	自動付与	理論的上限値	自動付与	理論的上限値
向上	10	12	18	18	16	15
変化なし	14	13	5	6	7	9
悪化	6	5	7	6	7	6

シア金融危機の原因は何か」（S230）や，知識を問う質問である「遺伝子組み換え食品とはどのようなものか」（S150），「組織犯罪対策三法案とはどのようなものか」（S160）などがある．

5.4 考察：課題ごとの失敗分析

提案手法により，ベースラインシステムに対するカバレッジが変化した課題数を**表 9** に示す．本節では，提案手法がベースラインよりカバレッジが低くなった場合について，失敗分析を行った．

事実報告型要約

失敗の原因は，ジャンル特性自動付与の誤り（S030），文タイプ自動付与の誤り（S160）があった．そのほか，識者意見タイプ（S100，S150，S180），個人性の記事（S140）の正解文がシステム要約から漏れた．

意見重視型要約

失敗の主な原因は，ジャンル特性自動付与の誤り（S040，S160），文タイプ自動付与の誤り（S110），両方の特性の自動付与の誤りの相乗効果（S030，S090）であった．そのほか，背景タイプ（S230，「ロシアは昨秋にも，アジア経済危機の影響で…外国投資家が市場から手を引き，金融危機に見舞われた.」），解説タイプ（S290，「国際宇宙ステーション計画の費用は開発・建設費だけで 4 兆円…とされている.」）の正解文がシステム要約から漏れた．

- これらの文のように，著者や識者の認識を含みつつ一般的な事実と取れる文を “著者意見” や “識者意見” タイプと判定することについては検討の余地がある．
- これらの文は，意見重視型要約では，“ロシア経済危機” や “国際宇宙ステーション建設” の問題点についての意見として参照要約に含まれている．現在の提案手法では，作成する要約の文脈を利用してこれらの文を抽出することはできない．

知識重視型要約

失敗の原因は，文タイプ自動付与の誤り（S040，S230）があった．そのほか，主記タイプ（S020，S210），背景タイプ（S240，S270），意見性の記事（S140）の正解文がシステム要約から漏れた．

ま と め

本実験では，人手で付与した文タイプとジャンル特性を利用して作成した要約のカバレッジを理論的上限値とすることにより，提案手法の有効性について検証した．提案手法の理論的上限値は，すべての情報のタイプについて有意に向上した．また，本手法の適用範囲を示すために，失敗分析を行った．理論的上限値との比較から，失敗の原因は，ジャンル特性と文タイプの自動付与誤りが 9 件，提案手法が観点を識別できていないものが 11 件となった．

この結果から，提案手法は利用者が重視する情報のタイプの識別に有効ではあるが，自動付与の精度に依存し，将来の課題として，要約の文脈を考慮して情報のタイプを識別する技術を開発すると改善が期待できることがわかった．

6. おわりに

本稿では，利用者の情報要求に適合する複数文書要約の手法として，重視する情報のタイプに焦点を当てた観点に基づく複数文書要約について検証した．事実報告，意見，知識の重視する情報のタイプを区別した実験用データセットを作成して評価した結果，文タイプとジャンル特性を手がかりとすることにより，参照要約に対するカバレッジが向上し，統計的有意差があった．また，3 種類の情報要求を区別した質問集合に対して要約を読んで質問にどれだけ答えられるか実験した結果，本提案手法は，重視する情報タイプに対応する質問に対する解答率が向上し，統計的有意差があった．これらの結果から，利用者の指定する要約の観点に応じた複数文書要約を作成する手法の有効性と有用性を示すことができた．

謝辞 この研究の一部は科学研究費補助金萌芽研究「検索意図と文書特性に基づいて特定の観点から内容をまとめる柔軟な複数文書自動要約」（課題番号 16650053）ならびに科学研究費補助金特定領域研究「不均質コンテンツに対する情報活用システムに関する研究」（課題番号 13224087）を受けて遂行された．また，研究の一部は栢森情報科学振興財団の助成を受けて遂行された．NTCIR-4 TSC の開催にあたってご

尽力された方々に対して感謝します.

参 考 文 献

1) Seki, Y., Eguchi, K. and Kando, N.: User-focused Multi-document Summarization with Paragraph Clustering and Sentence-type Filtering, *Proc.4th NTCIR Workshop on Research in Information Access Technologies: Information Retrieval, Question Answering, and Summarization*, pp.459–466 (2004).
2) Barry, C.L.: The Identification of User Criteria of Relevance and Documents Characteristics: Beyond the Topical Approach to Information Retrieval, Ph.D. Thesis, Syracuse University (1993).
3) Mani, I.: Automatic Summarization, *Natural Language Processing*, 1st edition, Vol.3, John Benjamins, Amsterdam, Philadelphia (2001).
4) 新村 出 (編)：広辞苑 第五版, 岩波書店 (2003).
5) 奥村 学, 難波英嗣：テキスト自動要約に関する研究動向（巻頭言に代えて), 自然言語処理, Vol.6, No.6, pp.1–26 (1999).
6) Tombros, A. and Sanderson, M.: Advantages of Query Biased Summaries in Information Retrieval, *Proc. 21st ACM SIGIR Conf. on Research and Development in Information Retrieval* (*SIGIR '98*), Melbourne, Australia, pp.2–10 (1998).
7) Mori, T., Nozawa, M. and Asada, Y.: Multi-Answer-Focused Multi-Document Summarization Using a Question-Answering Engine, *Proc. 20th Int'l Conf. on Computational Linguistics* (*COLING 2004*), Geneva, Switzerland, pp.439–445 (2004).
8) National Institute of Standars and Technology: Document Understanding Conferences (DUC) [online], *Document Understanding Conferences* (*DUC*) *website* (2001-2004). [cited 2004-10-26]. Avalilable from http://duc.nist.gov/.
9) Cardie, C., Wiebe, J., Wilson, T. and Litman, D.: Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering, *Proc. AAAI Spring Sympo. on New Directions in Question Answering*, Stanford, CA, pp.20–27 (2003).
10) Beineke, P., Hastie, T., Manning, C. and Vaithyanathan, S.: Exploring Sentiment Summarization, *Proc. AAAI Spring Sympo. on Exploring Attitude and Affect in Text: Theories and Applications* (*AAAI-EAAT 2004*), Stanford, CA, pp.12–15 (2004).
11) Pang, B. and Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proc. 42nd Ann. Meeting of the Assoc. for Computational Linguistics* (*ACL2004*), Barcelona, Spain, pp.271–278 (2004).
12) Xu, J., Weischedel, R. and Licuanan, A.: Evaluation of an Extraction-Based Approach to Answering Definitional Questions, *Proc. 27th ACM SIGIR Conf. on Research and Development in Information Retrieval* (*SIGIR 2004*), Sheffield, UK, pp.418–424 (2004).
13) Blair-Goldensohn, S., McKeown, K.R. and Schlaikjer, A.H.: Answering Definitional Questions: A Hybrid Approach, *New Directions in Question Answering*, Maybury, M.T. (Ed.), chapter 4, pp.47–57, AAAI/MIT Press (2004).
14) 藤井 敦, 渡邉まり子, 石川徹也：事典的 Web 検索サイトにおける複数文書要約の応用, 言語処理学会第 10 回年次大会発表論文集, pp.261–264, 言語処理学会 (2004).
15) Pomerantz, J.: Question Taxonomies for Digital Reference [online], Ph.D. Thesis, Syracuse University (2002). [cited 2004-10-26]. Avalilable from http://www.ils.unc.edu/~jpom/diss.html.
16) Seki, Y., Eguchi, K. and Kando, N.: Compact Summarization for Mobile Phones, *Mobile and Ubiquitous Information Access*, Crestani, F., Dunlop, M. and Mizzaro, S. (Eds.), Lecture Notes in Computer Science, Vol.2954, pp.172–186, Springer-Verlag, Heidelberg, Germany (2004).
17) McKnight, L. and Srinivasan, P.: Categorization of Sentence Types in Medical Abstracts, *Proc. American Medical Informatics Assoc.* (*AMIA*) *Sympo.*, Ottawa, Canada, pp.440–444 (2003).
18) Teufel, S. and Moens, M.: Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status, *Computational Linguistics*, Vol.28, No.4, pp.409–445 (2002).
19) Bazerman, C.: Speech Acts, Genres and Activity Systems: How Texts Organize Activity and People, *What Writing Does and How It Does It—An Introduction to Analyzing Texts and Textual Practices*, Bazerman, C. and Prior, P. (Eds.), pp.309–339, Lawrence Erlbaum Associates, Mahwah, NJ (2004).
20) Finn, A., Kushmerick, N. and Smyth, B.: Genre Classification and Domain Transfer for Information Filtering, *Advances in Information Retrieval, Proc. 24th BCS-IRSG European Colloquium on IR Research*, Crestani, F., Girolami, M. and van Rijsbergen, C.J. (Eds.), Lecture Notes in Computer Science, Vol.2291,

Springer-Verlag, Glasgow, UK, pp.353–362 (2002).
21) Endres-Niggemeyer, B.: *Summarizing Information*, Springer, Berlin (1998).
22) Biber, D., Conrad, S. and Reppen, R.: *Corpus Linguistics—Investigating Language Structure and Use*, Cambridge University Press (1998).
23) National Institute of Informatics: NTCIR (NII-NACSIS Test Collection for IR Systems) Project [online], *NTCIR* (*NII-NACSIS Test Collection for IR Systems*) *Project website* (1998-2004). [cited 2004-10-26]. Avalilable from http://research.nii.ac.jp/ntcir/.
24) Kando, N.: Overview of the Fourth NTCIR Workshop, *Proc. 4th NTCIR Workshop on Research in Information Access Technologies: Information Retrieval, Question Answering and Summarization*, National Institute of Informatics (2004).
25) Hirao, T., Okumura, M., Fukusima, T. and Nanba, H.: Text Summarization Challenge 3: Text Summarization Evaluation at NTCIR Workshop 4, *Proc. 4th NTCIR Workshop on Research in Information Access Technologies: Information Retrieval, Question Answering, and Summarization*, National Institute of Informatics (2004).
26) Hirao, T., Sasaki, Y. and Isozaki, H.: An Extrinsic Evaluation for Question-Biased Text Summariztion on QA tasks, *Proc. Workshop on Automatic Summarization at the Second Meeting of the North American Chapter of the Assoc. for Computational Linguistics* (*NAACL 2001*), Pittsburgh, PA, pp.61–68 (2001).
27) 神門典子：認識特性に基づくテキスト構造の分析，学術情報センター紀要，Vol.8, pp.107–126 (1996).
28) Cohen, J.: A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, Vol.20, No.1, pp.37–46 (1960).
29) Joachims, T.: *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*, Kluwer Academic Publishers (2002).
30) 工藤　拓：CaboCha/南瓜：Yet Another Japanese Dependency Structure Analyzer [online], 技術報告，奈良先端技術大学 (2004). [cited 2004-10-26]. Avalilable from http://chasen.org/~taku/software/cabocha.
31) 国立国語学研究所（編）：分類語彙表，国立国語学研究所資料集 14，増補改訂版 edition，大日本図書，東京 (2004).
32) Burman, P.: A comparative study of ordinary cross-validation, v-fold cross validationand the repeated learning-testing methods, *Biometrika*, Vol.76, No.3, pp.503–514 (1989).
33) Landis, J.R. and Koch, G.G.: The measurement of observer agreement for categorical data, *Biometrics*, Vol.33, pp.159–174 (1977).
34) Kessler, B., Nunberg, G. and Schutze, H.: Automatic Detection of Text Genre, *Proc. 35th Ann. Meeting of the Assoc. for Computational Linguistics joint with the 8th Conf. of the European Chapter of the Assoc. for Computational Linguistics* (*ACL/EACL '97*), Madrid, Spain, pp.32–38 (1997).
35) Karlgren, J. and Cutting, D.: Recognizing Text Genres with Simple Metrics Using Discriminant Analysis, *Proc. 15th Int'l Conf. on Computational Linguistics* (*COLING 1994*), Kyoto, Japan, pp.1071–1075 (1994).
36) Jing, H., McKeown, K., Barzilay, R. and Elhadad, M.: Summarization evaluation methods: Experiments and analysis, *American Assoication for Artificial Intelligence Spring Sympo. Series*, pp.60–68 (1998).

(平成 16 年 12 月 20 日受付)
(平成 17 年 4 月 6 日採録)

(担当編集委員　福島 俊一)

関　洋平（正会員）

1996 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．2005 年総合研究大学院大学情報学専攻博士後期課程修了．博士（情報学）．同年豊橋技術科学大学工学部情報工学系助手，現在に至る．自然言語処理，要約の研究に従事．ACM，ACL，電子情報通信学会，言語処理学会各会員．

江口　浩二（正会員）

1993 年同志社大学工学部電子工学科卒業．1999 年関西大学大学院工学研究科博士課程修了．博士（工学）．同年学術情報センター助手．2000 年国立情報学研究所助手，2002 年総合研究大学院大学助手を併任，現在に至る．2004 年フランス CLIPS-IMAG 研究所客員研究員，2005 年米国マサチューセッツ大学客員研究員．情報検索，Web テキスト処理等の研究に従事．ACM，電子情報通信学会，人工知能学会各会員．

神門 典子（正会員）

1994 年慶應義塾大学大学院文学研究科博士課程修了．博士（図書館・情報学）．同年学術情報センター助手．1995 年米国シラキウス大学情報学部客員研究員，1996〜1997 年デンマーク王立図書館情報大学客員研究員．1998 年学術情報センター助教授．2000 年国立情報学研究所助教授，2002 年より総合研究大学院大学助教授を併任，2004 年より国立情報学研究所教授ならびに総合研究大学院大学教授，現在に至る．テキスト構造を用いた検索と情報活用支援，言語横断検索，情報検索システムの評価等の研究に従事．ACM-SIGIR，ASIS&T，言語処理学会，日本図書館情報学会各会員．