

調停要約生成手法の改善と調停要約コーパスを用いた評価 Evaluation of the Method for Mediator Summary Generation Using the Mediator Summary Corpus

渋谷 英潔[†] 中野 正寛[‡] 石下 円香[†] 永井 隆広[‡] 森 辰則[†]

Hideyuki SHIBUKI Masahiro NAKANO Madoka ISHIOROSHI Takahiro NAGAI Tatsunori MORI

1. はじめに

Web上に存在する情報は、ブロードバンド化の進展やブログ等の普及に伴い、爆発的に増加し続けている。これらの情報の中には、出所が不確かな情報や利用者に不利益をもたらす情報などが含まれており、信頼できる情報を利用者が容易に得るための技術に対する要望が高まっている。しかしながら、情報の内容の真偽や正確性を検証することは困難である上に、その情報が意見などの主観を述べるものである場合には、利用者により考え方や受け止め方が異なることから、その真偽や正確性を検証することはさらに困難なものとなる。そのため、情報の信憑性は、最終的に個々の情報利用者が判断しなければならず、利用者による信憑性の判断を支援する技術の実現が優先して解決すべき課題であると考えられる。我々は、このような考えから、利用者の信憑性判断を支援する技術の実現に向けて研究を行っている [1-7]。

現在、ある情報の信憑性を Web のみを情報源として判断しようとした場合、Web 検索エンジンにより上位にランキングされた文書集合を読んで判断することが多い。しかしながら、例えば「ディーゼル車は環境に良いか?」というクエリで検索された文書集合には、「ディーゼル車は環境に良い」と主張する文書と「ディーゼル車は環境が悪い」と主張する文書の両方が含まれている場合があり、その対立関係をどのように読み解くべきかに関する手がかりを検索エンジンは示さない。ここでの対立関係の読み解き方とは、例えば、一方の内容が間違っているのか、それとも、両方の内容が正しく両立できるのか、といった点に関する可能性の示唆であり、もしも両立できるのであれば、何故対立しているように見えるのかに関する解説を提示することが望ましい。

我々は、検索された Web 文書の集合を、対立関係の読み解き方を簡潔に述べている文章に要約する調停要約の自動生成手法を提案している [1-4]。調停要約としての良さを評価する基準は一般的な目的的要約評価基準と異なるものであるため、ROUGE [8] に代表されるような既存の要約評価手法を適用して評価することは困難である。それゆえ、文献 [1-4] では、システムが Web から動的に生成した出力に対して人手で正解か否かを判断することで評価を行っていた。しかしながら、こうした主観的な評価方法では評価者による揺れの問題があり、手法の改善による効果を客観的に調査することは困難であった。このような背景から、我々は文献 [6] において、調停要約に関する評価・分析用の調停要約コーパスの作成を

行っており、本論文では、調停要約コーパスを利用して調停要約生成手法の客観的な評価を行う。また、ストップワードによる特徴語抽出の制限や逆説表現による対比構造の利用といった改善を行い、その効果の程度を明らかにする。

本論文の構成は以下の通りである。2. 章では、新しい要約概念である調停要約に関して説明する。3. 章では、関連研究について述べる。4. 章では、基本となる調停要約生成手法に関して説明する。5. 章では、ストップワードによる特徴語抽出の制限と逆説表現による対比構造の利用による改善手法を述べる。6. 章では、調停要約コーパスを用いた評価実験を行い、その結果を考察する。7. 章は結論である。

2. 調停要約

2.1 目的と定義

利用者がある情報の信憑性を疑うようになる原因の一つとして、その情報と対立するように見える情報が提示された場合がある。1. 章で述べたように、信憑性を判断したい言明¹ (以降、着目言明) をクエリとして検索した Web 文書集合には、着目言明を肯定する主張の文書と否定する主張の文書の両方が含まれていることが多い。しかしながら、互いに対立しているように見える言明の組の中には、好悪の評価や、一方が本当で他方が嘘であるという真に対立している言明の組も存在するが、互いが前提とする視点が異なるために実際には対立していない言明の組も存在する。例えば「ディーゼル車は環境に良い」と主張する文書を精読すると「CO₂の排出量が少ないので環境に良い」という文脈で述べられており、「ディーゼル車は環境が悪い」と主張する文書を精読すると「NO_xの排出量が多いので環境が悪い」という文脈で述べられているとする。この場合、前者は「地球温暖化」という観点から環境の良し悪しを述べているのに対して、後者は「大気汚染」という観点から述べており、互いの主張を否定する関係ではない。つまり、前提となる環境を明確にしない限り「ディーゼル車は環境に良いか?」というクエリが信憑性を判断できるような問いではないことを示しており、「あなたが想定している『環境』が地球温暖化を指しているなら環境に良いが、大気汚染を指しているならば環境が悪い」といった回答が、この例では適切であると考えられる。

我々は、このような一見対立しているように見えるが、実際はある条件や状況の下で互いの内容が両立できる関係を疑似対立、疑似対立が両立できるようになる条件や

[†]横浜国立大学大学院環境情報研究院, Graduate School of Environment and Information Sciences, Yokohama National University

[‡]横浜国立大学大学院環境情報学府, (ditto)

¹本論文では、主観的な意見や評価だけでなく、疑問の表明や客観的事実の記述を含めたテキスト情報を広く言明と呼ぶこととする。

着目言明: ディーゼル車は環境に良い

調停要約:

日本とアメリカでは、EUと違ってディーゼル車のイメージはよくありません。これは、環境に対する対策の違いだと思います。EUでは、呼吸器の病気を引き起こすNOxや浮遊微粒子(SPM)が多くても、地球温暖化の原因となる二酸化炭素(CO2)の少ないディーゼル車が環境にやさしい車として増えつつきているのです。

株式会社セラモ (<http://page.freett.com/ceramopower/qa.html>)

図1: 調停要約の例

状況を調停視点と定義し、疑似対立を読み解くための手掛かりとなる簡潔な文章を提示することで利用者の信憑性判断を支援することを目的としている。そして、着目言明をクエリとして検索したWeb文書集合から、疑似対立を読み解くための手掛かりとなる記述を簡潔な文章として取り出したものを調停要約と定義する。

2.2 基本的な考え方

情報信憑性判断を支援するという目的から、システムの要約過程において信憑性が損なわれる危険性を可能な限り減らす必要がある。そのためには、元文書中の記述をそのまま抜粋して利用者に提示することが望ましい。我々は、文の連続をパッセージとして定義し、パッセージ抽出に基づいて調停要約の生成を実現する²。

図1は、「ディーゼル車は環境に良い」を着目言明とした場合の理想的な調停要約の例である。調停要約のボックス内の記述は、図中のURLで示す会社のサイトから人手により抽出したものである。斜字体で示した部分には「環境に対する対策の違い」という調停視点がかかれている。Web上には、こういった対立関係が両立可能であることを示した記述が存在していることがあり、そのような記述をパッセージ単位で抜粋して提示するという方法が調停要約の基本的な考え方である。

2.3 観察される特徴

図1のような調停要約を生成するために、我々は、文献[6, 7]等において、情報信憑性判断のための要約コーパスを手で作成し分析を行った。その結果、調停要約として相応しい記述には以下の3つの特徴がある場合が多いことが観察された。

第一の特徴は、着目言明と関連性が高いことである。調停要約は着目言明に対する回答であるため、適切な調停要約には高い関連性が必要であると考えられる。第二の特徴は、公平性が高い、すなわち、着目言明を肯定する意見や根拠等と否定する意見や根拠等の両方に等しく言及していることである。調停要約は、疑似対立を読み解くための手掛かりとなる記述であるため、肯定側と否定側の双方の主張が含まれる記述がより適切であると考えられる。第三の特徴は、上記の特徴を示す単語や表現が密集していることである。高い密集度をもつパッセージは、簡潔な要約としての妥当性に加えて、両方の意見や根拠等を対比的に記述していることが多く、調停要約

²文献[5]において、1つのパッセージで両立可能となる状況を明示的に説明する直接調停要約と、状況の一部を説明するパッセージを複数組み合わせることで状況の全体を暗に示す間接調停要約の2種類を定義しているが、本論文では直接調停要約を対象としており、以後、直接調停要約を単に調停要約と記す。

としてより適切であると考えられる。4章の手法は、これらの特徴を用いることで調停要約に相応しいパッセージの抽出を行っている。

3. 関連研究

利用者の情報信憑性の判断を支援する技術には幾つかのアプローチが考えられる。まず、利用者が着目する話題や言明に関連するWeb文書に対して、対立の構図や根拠関係などを多角的に俯瞰することを支援する技術がある。Akamine et al.[9]は、利用者が入力した分析対象トピックに関連するWebページに対して、主要・対立表現を俯瞰的に提示するシステムWISDOMを開発している。Murakami et al.[10]は、Web上に存在するさまざまなテキスト情報について、それらの間に暗に示されている同意、対立、弱い対立、根拠などの意味的關係を解析する言論マップの生成課題を論じている。藤井[11]は、Web上の主観情報を集約し、賛否両論が対立する構図を論点に基づいて可視化している。これらの研究では、対立関係にある記述の提示に焦点があり、提示された対立関係の読み解き方に関しては対象としていない。対立関係の把握が容易になるような要約を利用者に提示できれば、着目言明に関連する話題や言明群の全体像が把握しやすくなると考えられる。それゆえ、これらの技術とは相補的な関係であるといえる。

Finn et al.[12]は、Web上の新聞記事を対象として、コラム等の主観的な記事と事実を伝える客観的な記事に分類する研究を行っている。好悪といった主観に依存する言明間の対立関係の場合、互いの内容は両立することができるため、主観的であるか否かの情報は対立関係の読み解き方に役立つと考えられる。しかしながら、ディーゼル車の例のように客観的な内容の対立関係においても疑似対立となる場合があり、そのような場合の読み解き方に調停要約は焦点を絞っている。

他にも、利用者が着目する言明に対するWeb上の意見の変遷と意見が変わった要因を提示する河合らの研究[13]、Webページのレイアウト情報を利用して情報発信者の名称を抽出するMiyazaki et al.の研究[14]、Webページの情報発信構成の考え方に基いて情報発信者の同定を行う加藤らの研究[15]等がある。これらの研究は、発信された情報の内容ではなく、「いつ」「誰が」発信したかといった面から利用者の判断を支援するアプローチを取っている。したがって、例えば、図1のように調停要約と並べて情報発信者を提示することで、さらに利用者への支援が容易になると考えられる。

調停要約は、複数文書を対象とした抜粋型の報知的要約の一つである。その中でも、橋本ら[16]の研究のような、要約対象文書群から「まとめ文章」を取り出すことにより要約する手法に属する。文献[16]では、新聞記事を対象にしており、複数文書の記述内容に齟齬があることは想定せずに、複数記事の内容をそのままとめることが目的である。一方で調停要約では、まず、様々な立場の人物や組織が互いに対立した主張をしているようにみえる記述を含む文書集合を要約対象にしている点が異なる。さらに、得られる要約の中に、疑似対立とその読み解き方が含まれるようにすることで、情報信憑性の判断に寄与することを目的としている。

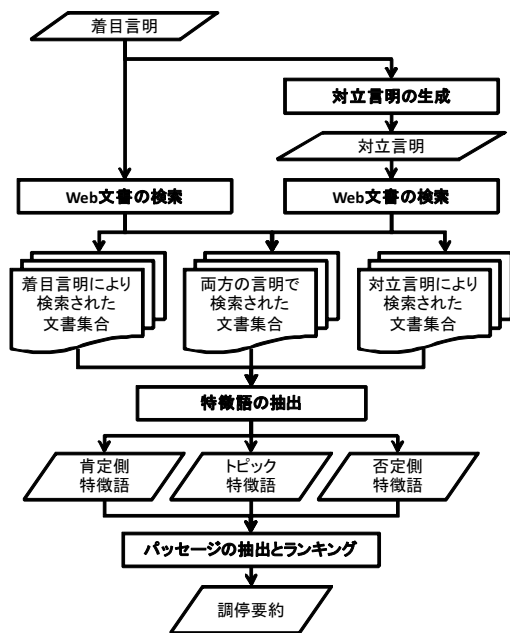


図2: 調停要約生成手法の全体の流れ

利用者が入力したクエリに対して簡潔な文章を出力するという枠組みは、Non-Factoid型の質問応答システム[17]と類似している。質問応答として捉えると、着目言明を入力としてその真偽を問うYes/No型の質問応答となることが考えられるが、調停要約の場合には、単純にYes/Noで回答できる質問ではないということを気付かせる文章を出力するという点で質問応答の考え方とは異なっている。したがって、質問応答システムにおいてYesとNoの両方の解が得られるような場合に調停要約を提示するといった利用が考えられるが、本論文では、質問応答システムとの連携は今後の課題として、単純にYes/Noでは回答できない質問が入力されることを前提としている。

4. 調停要約生成手法

4.1 全体の流れ

我々は、2.3節で述べた特徴をもつパッセージを調停要約として出力するために、以下の3種類の特徴語に基づく手法を文献[1, 2]において提案している。本章では、この手法を説明する。

まず、着目言明との関連度を近似的に求めるために、着目言明のトピックとの関連性が高い語をトピック特徴語として定義する。本論文では、着目言明中の内容語をトピック特徴語とした³。次に、着目言明を肯定または否定する意見や根拠等に言及しているかを近似的に求めるために、双方の意見や根拠に現れやすい単語を肯定側特徴語および否定側特徴語としてそれぞれ定義する。肯定側特徴語と否定側特徴語は対となる語であると考えられるが、着目言明中に含まれているとは限らず、また、必ずしも一般的な対義語に限定されていない⁴。それゆ

³トピック特徴語は理論的に着目言明中の単語に限定されるものではない。着目言明以外の語句への拡張は今後の課題である。

⁴例えば、図1中の「CO₂」と「NO_x」、「EU」と「日本」といった単語の組も、「ディーゼル車は環境に良い」という着目言明において

え、これらは着目言明だけでは決定できないことが多い。我々は、Webから着目言明に肯定的な文書集合と否定的な文書集合を検索し、各々の文書集合に現れやすい単語を抽出することで、肯定側特徴語と否定側特徴語を求める。

図2に全体の流れを示す。入力として着目言明が与えられると、まず、着目言明と反対の極性をもつ対立言明の生成を行う。次に、着目言明と対立言明を用いてWeb文書集合の検索を行う。検索された文書集合は、肯定側および否定側の特徴語を抽出するために用いる。抽出されたトピック関連語、肯定側特徴語、否定側特徴語を用いて、2.3節で述べた特徴に基づき、検索されたWeb文書中の各文の直接調停としてのスコアを計算し、スコアに応じてパッセージを抽出、ランキングして出力する。次節以降で、上記の各段階を詳細に説明する。

4.2 対立言明の生成とWeb文書の検索

対立言明は、対義語辞書を用いて、着目言明中の1語をその対義語で置換することにより生成する。また、着目言明中の置換された語を肯定側特徴語、置換された語に対応する対立言明中の語を否定側特徴語として保持する。置換可能な単語が存在しない場合、対立言明は生成しない。

次に、着目言明と生成された対立言明をクエリとして、自然文をクエリとして使用できるTSUBAKI[18]によりWeb文書の検索を行う。検索された文書は以下の3種類の文書集合に分類する。第一の集合 D_{que} は、着目言明により検索されたが対立言明により検索されなかった文書の集合、第二の集合 D_{inv} は、対立言明により検索されたが着目言明により検索されなかった文書の集合、第三の集合 D_{both} は、着目言明をクエリとする検索と対立言明をクエリとする検索の両方で得られた文書の集合である。

TSUBAKIには、クエリ中の最後の用言について、その反義・否定表現を加えて検索する機能がある。この機能を有効にした状態で着目言明により検索された文書も、対立言明により検索された文書とみなすこととした。また同様に、その機能を有効にした状態で対立言明により検索された文書も、着目言明により検索された文書とみなすこととした。これにより、対義語辞書で対立言明が生成されなかった場合でも、着目言明と対立する内容を含む文書を検索することができる。

4.3 特徴語の抽出

肯定側特徴語と否定側特徴語は、検索されたWeb文書から抽出する。最初に、単語 w における肯定側スコア $sc_{POS}(w)$ と否定側スコア $sc_{NEG}(w)$ を、式(1)と式(2)によりそれぞれ計算する。

$$sc_{POS}(w) = \frac{df(w, D_{que}) \cdot tf(w)}{df(w, D_{inv}) + 1} \quad (1)$$

$$sc_{NEG}(w) = \frac{df(w, D_{inv}) \cdot tf(w)}{df(w, D_{que}) + 1} \quad (2)$$

ここで $tf(w)$ は検索された全ての文書における単語 w の頻度であり、 $df(w, D)$ は文書集合 D において w を含んでいる文書の数である。

は対となる単語とみなすことができる。

表 1: 抽出された特徴語の例

単語	$rank_{tf}$	$rank_{POS}$	$rank_{NEG}$	特徴語
レーシック	1	1	1	非特徴語
手術	2	2	2	非特徴語
視力	3	3	3	非特徴語
検査	19	13	60	肯定側
メガネ	47	58	75	非特徴語
失明	61	5,045	20	否定側
効果	66	72	111	肯定側
合併症	77	206	58	否定側

D_{que} には多く出現するが、 D_{inv} には出現しない単語は、式(1)により、 $sc_{POS}(w)$ が高くなる。これにより、肯定側特徴語の候補となる語を上位にランキングできると考えられる。否定側特徴語に関しても同様に、式(2)により上位にランキングされた語を候補とできると考えられる。

しかしながら、検索された全文書に満遍なく出現する語は $tf(w)$ が高いので $sc_{POS}(w)$ と $sc_{NEG}(w)$ の両方が高い値となりやすい。このような語は、着目言明との関連性が高い可能性があるが、着目言明を肯定または否定するのに用いられやすいと考えられない。それゆえ、肯定側特徴語や否定側特徴語として抽出すべきではない。また、文書集合 D_{que} と D_{inv} の大きさが同じであるとは限らないため、抽出には $sc_{POS}(w)$ と $sc_{NEG}(w)$ の値にそれぞれ基づいた順位 $rank_{POS}(w)$ と $rank_{NEG}(w)$ を用いることとした。まず、全ての単語を $tf(w)$ の値が大きい順にランキングし、上位 C_{rank} 語を抽出する特徴語の候補とする。その候補の中で、 $rank_{POS}(w)$ の方が $rank_{NEG}(w)$ よりも閾値 C_{dif} を超えて大きい場合には肯定側特徴語、 $rank_{NEG}(w)$ の方が $rank_{POS}(w)$ よりも C_{dif} を超えて大きい場合には否定側特徴語とし、その他の単語は特徴語とみなさないようにした。表1に、着目言明が「レーシック手術の安全性は高い」の場合に抽出された特徴語の例を示す。最後に、4.2節で対立言明を生成する際に保持した肯定側および否定側の特徴語集合を、抽出された肯定側および否定側の特徴語集合にそれぞれ加える。

トピック特徴語集合は、肯定側および否定側の特徴語集合に含まれない着目言明中の単語の集合とした。

4.4 パッセージの抽出とランキング

調停要約となるパッセージは以下の処理で要約対象文書群から抽出しランキングする。

第一段階の処理では、調停要約として適切ではない文を認識する。本論文では、1以上の動詞と2以上の名詞、または、2以上の動詞と1以上の名詞をもつ文を十分な文と定義し、これを満たさない文を不十分な文とした。また、文末が「...」という省略表現で終わっている文を省略された文とし、以上の2種類を調停要約として適切ではない文とした。不十分かつ省略された文の場合には、単に省略された文とした。これらの文は、次の処理である文スコアの計算でペナルティを受ける。

第二段階の処理では、文のスコアを計算する。 KW を、トピック特徴語、肯定側特徴語、否定側特徴語の和集合とした場合、文 s の基本スコア $sc_{BS}(s)$ を式(3)によって計算する。

$$sc_{BS}(s) = \frac{\sum_{w \in KW} appear(w, s)}{|KW|} \quad (3)$$

ここで $appear(w, s)$ は単語 w が文 s 中に現れていれば1、そうでなければ0を返す関数である。次に、文 s が2.3節で述べた公平性の特徴をもつ場合のボーナス $R_S(s)$ と、調停要約として適切ではない文の場合のペナルティ $P_S(s)$ を以下の式(4)と式(5)に従って計算する。

$$R_S(s) = \begin{cases} C_{eith} & (s \text{ がトピック特徴語を含む} \\ & \text{2種類の特徴語に言及して} \\ & \text{いるとき)} \\ C_{both} & (s \text{ が全種類の特徴語に言及} \\ & \text{しているとき)} \\ 1 & (\text{otherwise}) \end{cases} \quad (4)$$

$$P_S(s) = \begin{cases} C_{ins} & (s \text{ が不十分な文のとき)} \\ C_{omit} & (s \text{ が省略された文のとき)} \\ 1 & (\text{otherwise}) \end{cases} \quad (5)$$

C_{eith} は1より大きい定数、 C_{both} は C_{eith} より大きい定数、 C_{ins} と C_{omit} は1より小さい定数である。 C_{eith} の条件は、文 s がトピック特徴語に言及している⁵ことに加えて、肯定側特徴語または否定側特徴語のどちらかに言及している場合を指し、 C_{both} の条件は、文 s がトピック特徴語、肯定側特徴語、否定側特徴語の全てに言及している場合を指す、式(6)に従って、基本スコア $sc_{BS}(s)$ に $R_S(s)$ と $P_S(s)$ を乗じた値が文 s のスコア $sc_S(s)$ となる。

$$sc_S(s) = sc_{BS}(s)R_S(s)P_S(s) \quad (6)$$

第三段階の処理では、パッセージの断片化を防ぐために文スコアを平滑化する。平滑化には窓幅 L のハン窓を用いることとし、文書中の i 番目の文の平滑化スコア $sc_M(s_i)$ を式(7)により計算する。

$$sc_M(s_i) = sc_{BM}(s_i)R_M(i, \frac{L}{2})P_M(s_i) \quad (7)$$

$$sc_{BM}(s_i) = \sum_{j=-\frac{L}{2}}^{\frac{L}{2}} (sc_S(s_{i+j}) \cdot hf(j)) \quad (8)$$

$$hf(k) = 0.5 + 0.5 \cos 2\pi \frac{k}{L} \quad (9)$$

$$R_M(i, l) = \begin{cases} C_{smo} & (s_{i \pm l} \text{ の範囲で全種類の特} \\ & \text{特徴語に言及しているとき)} \\ 1 & (\text{otherwise}) \end{cases} \quad (10)$$

$$P_M(s) = \begin{cases} C_{omit} & (s \text{ が省略された文のとき)} \\ 1 & (\text{otherwise}) \end{cases} \quad (11)$$

⁵本論文での「特徴語に言及している」とは、例えば、複数ある特徴語の内の少なくとも1語が現れている状態のことを意味する。

表 2: ストップワードの一覧

分類	表現
著作権表示	copyright, (c), all rights reserved
ブログ, ニュースサイト	特集, スコア, カテゴリ, category, 画像, 写真, 記事, マップ, ニュース, 新規, フォーム, 登録, http, www, ウェブ, インターネット, ネット, サイト, ブログ, blog, weblog
投稿サイト	投稿, posted, 登録, 日時, 質問, 解決, 解答, 回答, 解決
通販, オークション	広告, 商品, 製品, 販売, 発売, 発送, 時期, 定価, 価格, 税込, 送料, 無料, 市場, ショッピング, 比較, オークション, 違反, 報告, PR
サイト内機能	リンク, top, トップ, ログイン, ログアウト, ページ, ホーム, 前, 戻る, 続き, 読む, more, コメント, トラックバック, バック, トラバ, comment, trackback, tracked, 一覧, 全文, 編集
アフィリエイト, アクセス対策	人気, ランキング, アクセス, アップ, ranking, RSS
サイト名	wikipedia, yahoo, 楽天, はてな
前置詞	by, on, at
その他	問い合わせ, 書く, ID, 検索, description, 転載, vol, PDF ファイル, 関連, キーワード, 参考, 文献, ヘルプ, 利用, 方法, 規約, プライバシーポリシー, 更新, 設定, メール, 笑 (1文字のアルファベット) (1文字の記号)

C_{smo} は1より大きい定数であり, $R_M(i, l)$ は, 文 s_i を中心とした長さ l 文の文列が, トピック特徴語, 肯定側特徴語, 否定側特徴語の全てに言及している場合の, 公平性の観点によるボーナスである. P_M は, 省略された文のスコアを平滑化する際のペナルティである. 不十分な文は, パッセージ中の前後の文と適切につながることで, 読み手に有益な情報を伝えることができるかもしれないが, 省略された文の場合は基本的にそうならない. それゆえ, 不十分な文のスコアには平滑化の際にペナルティを与えず, 省略された文の場合にのみペナルティを与える.

第四段階の処理では, パッセージの抽出とランキングを行う. 文書中における $sc_M(s)$ の最大値の $\frac{1}{C_{seg}}$ 以上の $sc_M(s)$ をもつ文の連続を1つのパッセージとして抽出する. C_{seg} は, 抽出するパッセージサイズに影響する定数で, C_{seg} の値が大きいほど大きなパッセージとなりやすい. パッセージ p のスコア $sc_P(p)$ は, 以下の式 (12) により計算する.

$$sc_P(p) = R_P(p) \max_{s \in p} sc_M(s) \quad (12)$$

$$R_P(p) = \begin{cases} C_{pas} & (p \text{ が全種類の特徴語に} \\ & \text{言及しているとき)} \\ 1 & (otherwise) \end{cases} \quad (13)$$

C_{pas} は1より大きい定数であり, $R_P(p)$ は, p 中の文列が, トピック特徴語, 肯定側特徴語, 否定側特徴語の全てに言及している場合の, 公平性の観点によるボーナスである. また, パッセージの長さは, 簡潔性という観点からは短い方がよいが, 調停要約の性質上, ある程度の長さは必要になると考えられる. それゆえ, 抽出されたパッセージの長さが目標とする長さ C_{len} に近いほど上位にランキングされるべきである. 以上から, パッセージの最終スコア $sc_F(p)$ は式 (14) に従って計算される.

$$sc_F(p) = \exp(sc_P(p) - C_{err} \cdot E(p)) \quad (14)$$

表 3: 逆接表現の一覧

分類	表現
接続詞	だが, しかし, しかしながら, それでいて, けれど, けれども, ところが, でも, 一方
接続助詞	けれど
連語	ではなく, 逆に, 対し, 反面, 正反対, だけでなく

$$E(p) = |C_{len} - nc(p)| \quad (15)$$

$nc(p)$ はパッセージ p の文字長, C_{err} は目標とする長さとの誤差の影響に関する係数である.

5. 調停要約生成手法の改善

4.章で述べた手法に対し, ストップワードによる特徴語抽出の制限と, 逆接表現による対比構造の利用による改善を行い, 精度向上を試みる.

5.1 ストップワードによる特徴語抽出の制限

我々の手法の精度は, 抽出される肯定側特徴語と否定側特徴語の精度によるところが大きい. 文献 [1] の実験結果から, 4.3 節で述べた方法では, 着目言明の内容と無関係な要因による頻度の偏りに対処することができないことが分かった. 着目言明の内容と無関係な要因とは, 検索された文書集合に通販サイトが多く含まれているといった事柄であり, 無関係な要因による頻度の偏りの例としては, 「マイナスイオンは健康に良い」という着目言明において, 肯定側の文書集合に通販サイトが多くを占める一方で否定側の文書集合には殆ど存在しないといったものが挙げられる. このような場合, 通販サイトに頻出する「税込」や「送料」といった語が特徴語として抽

表 4: 着目言明ごとのパッセージ数と各手法の平均精度

着目言明	総数	正解	BL	SW	AC	SWAC	SW _{OR}
コラーゲンは肌に良い	7,392	32	0.0071	0.0075	0.0123	0.0131	0.0072
飲酒は健康に良い	8,218	93	0.0546	0.0727	0.0823	0.1155	0.0626
炭酸飲料はからだに悪い	8,330	50	0.1316	0.1556	0.2549	0.2848	0.1260
原発は地震でも安全である	8,092	57	0.0393	0.0441	0.0584	0.0590	0.0401
車内での携帯電話の使用は控えるべきである	9,111	48	0.1216	0.1225	0.0543	0.0524	0.0968
嘘をつくのは悪いことである	7,813	36	0.0724	0.1642	0.1289	0.2219	0.0329

出されてしまい、結果として調停要約の精度に悪影響を及ぼすと考えられる。

藤井の研究 [11] における論点の抽出においても同様の問題が生じており、人手で設定したストップワードを用いることで対処している。我々も、文献 [1] の実験結果⁶を分析することで、表 2 に示すストップワードを設定した。ただし、表中の分類は厳密なものではなく、ストップワードの大雑把な使われ方を示したものである。4.3 節の処理において、抽出された肯定側および否定側の特徴語がストップワードであった場合には、最終的な肯定側および否定側の特徴語集合から除外する。

5.2 逆接表現による対比構造の利用

調停要約はその性質上、「NO_x は排出するが CO₂ は排出しない」や「EU でのイメージは良いが日本でのイメージは悪い」といった対比構造を含むことが多いと考えられる。そこで、対比構造を見つける手掛かりの一つとして逆接表現を利用する。逆接表現を含むパッセージは、対比構造が存在するパッセージである可能性が高いと考えられるため、逆接表現を含むパッセージを、含まないパッセージよりも上位にランキングする。本論文では、文献 [4] で用いたものと同じ表 3 に示す逆接表現を用いた。なお、表 3 の分類は、MeCab⁷のデフォルトの品詞体系に基づいている。

本論文では、4. 章の手法の後処理として、逆接表現を含まないパッセージを再ランキングするという方法で実装する。4. 章の手法でランキングされたパッセージ集合を入力として、上位パッセージから順に逆接表現を含んでいるか判定し、逆接表現を含まないパッセージを最下位に再ランキングする。

6. 評価実験

6.1 目的と手順

本実験の目的は、これまで主観的な評価しか行われてこなかった調停要約生成手法に対して、5. 章の改善手法の有効性を客観的指標で示すことである。調停要約生成手法およびその改善手法を客観的に評価するために、文献 [6] で作成した調停要約コーパスを用いる。調停要約コーパスには、「コラーゲンは肌に良い」、「飲酒は健康に良い」、「炭酸飲料はからだに悪い」、「原発は地震でも安全である」、「車内での携帯電話の使用は控えるべきであ

表 5: パラメータの値

C_{ins}	0	C_{omit}	0
C_{eith}	2	C_{both}	3
C_{smo}	2	C_{pas}	3
C_{rank}	100	C_{dif}	20
C_{len}	300	C_{err}	0.02
L	5	C_{seg}	3

る」、「嘘をつくのは悪いことである」の 6 つの着目言明に対して人手で作成した調停要約が収録されている。各着目言明に対して 4 名の作業者が割り当てられ、各作業者は 500 程度の Web 文書集合から調停要約として適切なパッセージ集合を抽出している⁸。

調停要約コーパスから以下の手順で実験データを作成した。まず、作業者により調停要約として抽出されたパッセージ集合を正解のパッセージ集合とする。コーパスに収録されている Web 文書には、調停要約となるパッセージ以外のパッセージ境界がないため、正解パッセージ集合の平均文字長を計算し、各文書の先頭から平均文字長 $-\alpha$ を超えた文境界で分割する。このとき、正解パッセージを含む文書を分割対象外として、分割されたパッセージ集合を不正解のパッセージ集合とした。正解および不正解のパッセージ集合の中には、平均文字長との差が極めて大きいものがあるため、平均文字長 $\pm\alpha$ の範囲にない長さのパッセージを実験データから除外した。平均文字長は 230 であり、 α の値は経験的に 30 とした。着目言明ごとの総パッセージ数と正解パッセージ数は、表 4 の総数と正解に示す値となった。

4. 章では、検索された文書を要約対象文書として説明したが、本実験では上記の実験データを要約対象文書とするため、検索された文書は特徴語の抽出のためだけに用いられる。実験データは既にパッセージ単位に切り出されているため、4.4 節の四段階目で行うパッセージの抽出は行わずランキングのみを行う。また、4. 章で述べた各パラメータの値は、各パラメータの性質から暫定的に表 5 の値に設定した。

⁸調停要約コーパス自身の客観性に関する厳密な議論は今後の課題であるが、文献 [6] での分析結果より、主要な調停視点に関しては作業員間で共通していることが明らかになっている。そのため、客観的評価のために調停要約コーパスを用いることに大きな問題は生じないと考えている。

⁶文献 [1] の実験で用いた着目言明は、6. 章の実験で用いる着目言明と異なるものである。

⁷<http://mecab.sourceforge.net/>

表 6: 上位 r 件の適合率

上位 r 件	BL	SW	AC	SWAC	SW _{OR}
10	0.100	0.267	0.233	0.300	0.083
20	0.142	0.158	0.175	0.192	0.133
30	0.122	0.139	0.183	0.228	0.100
100	0.112	0.113	0.133	0.133	0.088
1,000	0.028	0.029	0.026	0.026	0.026

表 7: 上位 r 件の再現率

上位 r 件	BL	SW	AC	SWAC	SW _{OR}
10	0.019	0.051	0.044	0.057	0.016
20	0.054	0.060	0.066	0.073	0.051
30	0.070	0.079	0.104	0.130	0.057
100	0.212	0.215	0.253	0.253	0.168
1,000	0.535	0.551	0.497	0.497	0.509

評価指標には適合率と再現率を用いた。調停要約として適切なパッセージは、コーパスに収録されている Web 文書集合から網羅的に抽出しているため、コーパス中の Web 文書を要約対象文書として処理を行うと再現率を計算することができる。また、調停要約として適切なパッセージ群が可能な限り多く上位にランキングされているか調査するために、TREC⁹や NTCIR¹⁰の情報検索タスクで広く用いられている平均精度を用いた。第 r 位の適合率 $Pre(r)$ と再現率 $Rec(r)$ 、および平均精度 AP は、それぞれ以下の式により計算される。

$$Pre(r) = \frac{correct(r)}{r} \quad (16)$$

$$Rec(r) = \frac{correct(r)}{R} \quad (17)$$

$$AP = \frac{1}{R} \sum_r I(r) Pre(r) \quad (18)$$

R は正解パッセージの総数、 $correct(r)$ は第 r 位までの出力パッセージに含まれる正解パッセージ数、 $I(r)$ は第 r 位のパッセージが正解ならば 1、不正解ならば 0 を返す関数である。

6.2 結果と考察

ベースラインとなる 4. 章で説明した手法を BL、表 2 に示すストップワードによる特徴語抽出の制限を加えた手法を SW、逆接表現による対比構造を利用した手法を AC、両方の改善を加えた手法を SWAC として、着目言明ごとの平均精度を表 4 に示す。また、上位 10 件、20 件、30 件、100 件、1,000 件の場合の適合率と再現率を表 6 と表 7 にそれぞれ示す。表 6 と表 7 の値は全ての着目言明の平均値である。さらに、表 2 に示すストップワードの有効性を調査するために、表 2 のストップワ

ドの代わりに藤井の研究 [11] で用いているストップワードを用いて特徴語抽出の制限を加えた手法 SW_{OR} の結果も各表に示す。

表 4 の結果から、着目言明ごとのばらつきはあるものの、ストップワードや逆接表現を利用することで全体的に精度を改善できたことを確認した。また、ストップワードと逆接表現の効果は互いに補う関係にあり、両者を組み合わせた場合に最も精度が改善できることが示された。しかしながら、着目言明が「車内での携帯電話の使用は控えるべきである」の場合には、ストップワードを用いた SW では精度が向上したが、逆接表現を用いた AC および SWAC では BL よりも低下した。逆接表現を用いたことで精度が低下した原因として、「実験的には、距離が 15 cm 以上の間隔ならば安全のようである」のような逆接表現を伴わずに両立可能となる条件を記述している正解パッセージを下位にランキングしたことが挙げられる。この問題を解決するためには、直接ランキングを操作するのではなく、逆接表現を含まないパッセージのスコアに対してペナルティを与えるなどの非決定的な処理を行う必要があると考えられる。また、逆接表現以外の両立可能となる条件の記述スタイルを利用することも考えられる。これらは今後の課題である。

表 2 に示すストップワードの利用は、全ての着目言明において改善効果があった。一方で、文献 [11] のストップワードを用いた SW_{OR} の値は、全ての着目言明で SW よりも低く、幾つかの着目言明では BL よりも低下した。したがって、文献 [11] のストップワードは、対立の論点抽出のためのものであるが、調停要約における特徴語抽出には必ずしも適していないことが分かった。今回の実験で表 2 のストップワードの有効性が示されたが、最善のものであるとは考えていない。今後、より多くの着目言明の結果に基づいてストップワードのリストを改善していきたいと考えている。

次に、表 6 と表 7 の結果から、各手法の改善効果がどのランクのパッセージに及んだかを考察する。表 4 の SW_{OR} の平均精度は、基本的に BL よりも高い値を示しているのに対し、表 6 と表 7 では、上位 1,000 件までの全ての値で BL よりも低かったことが分かる。このことは、SW_{OR} の改善効果が 1,000 位以下の下位パッセージにしか及ばなかったことを示している。一方、SW、AC、SWAC の改善効果は正解パッセージを 100 位以上に引き上げており、特に上位 10 位における改善効果が大きい。利用者が読むことができる調停要約の数は、労力的に限られていることから、可能な限り上位に正解パッセージが集まっていることが望ましい。したがって、上位パッセージの改善効果が大きいことから、5. 章で述べた改善手法が有効であったと言える。

ストップワードと逆接表現を用いることによる有効性は示されたが、改善の余地は多く残されている。表 6 から上位 10 件の適合率は 3 割程度しかなく、表 7 の再現率から 1,000 位以下に約半数の正解パッセージが埋もれていることが分かる¹¹。本論文では、ストップワードを

¹¹ただし、文献 [19] での 112 名を対象とした実証実験において、対立関係の読み解き方を示すという調停要約のタスクは、既存手法では達成困難なタスクであることが示されており、改善前の精度であっても既存手法を大きく上回る評価を得られたことが示されている。

⁹<http://trec.nist.gov/>

¹⁰<http://research.nii.ac.jp/ntcir/index-ja.html>

特徴語抽出にしか用いていないが、「copyright」や「トップへ戻る」などを含むパッセージは調停要約として不適切である可能性が高いため、パッセージの調停要約としての適切性を測る指標にストップワードを利用することで適合率を向上させることが考えられる。また、「は大きく は小さい」といった対義語の組により対比構造を示すなど、逆接表現の他にも対比構造を示す手掛かりは存在するため、そういった手掛かりを用いることで再現率を向上させることが考えられる。

7. おわりに

本論文では、調停要約コーパスを用いた調停要約生成手法の客観的な評価を行った。また、ストップワードによる特徴語抽出の制限や逆説表現による対比構造の利用等の改善を行い、その効果の程度を明らかにした。これまで、既存の要約評価基準と異なる基準をもつ調停要約の評価は、システムが Web から動的に作成した出力に対して主観的評価を行っていたため、評価者による揺れの問題があり、手法の改善による効果を客観的に調査することが困難であった。今回、調停要約コーパスを用いた客観的な評価が可能となったことで、改善の効果を客観的に調査することが容易になり、今後の改善に向けた環境を整えることができた。

今後の課題として、パッセージのスコア計算の改善、ストップワードや逆接表現のリストの拡充、逆接表現以外の対比構造の手掛かり表現の調査、各パラメータの最適値の調査等を行う予定である。

謝辞

本研究の一部は、科学研究費補助金 (No.22500124) の助成を受けたものである。

参考文献

- [1] H. Shibuki, T. Nagai, M. Nakano, R. Miyazaki, M. Ishioroshi and T. Mori, "A Method for Automatically Generating a Mediator Summary to Verify Credibility of Information on the Web," *Proc. of the 23rd International Conference on Computational Linguistics* (2010).
- [2] 中野 正寛, 渋木 英潔, 宮崎 林太郎, 石下 円香, 永井 隆広, 森 辰則, "直接調停要約自動生成システム HERMeS の実装と性能評価," 言語処理学会第 17 回年次大会発表論文集 (2011).
- [3] 石下 円香, 渋木 英潔, 中野 正寛, 宮崎 林太郎, 永井 隆広, 森 辰則, "直接調停要約自動生成システム HERMeS の言論マップとの連携," 言語処理学会第 17 回年次大会発表論文集 (2011).
- [4] 永井 隆広, 渋木 英潔, 中野 正寛, 石下 円香, 宮崎 林太郎, 森 辰則, "直接調停要約自動生成システム HERMeS の対比表現を用いた精度向上," 言語処理学会第 17 回年次大会発表論文集 (2011).
- [5] K. Kaneko, H. Shibuki, M. Nakano, R. Miyazaki, M. Ishioroshi and T. Mori, "Mediator Summary Generation: Summary-Passage Extraction for Information Credibility on the Web," *Proc. of the 23rd Pacific Asia Conference on Language, Information and Computation* (2009).
- [6] 渋木 英潔, 中野 正寛, 宮崎 林太郎, 石下 円香, 永井 隆広, 森 辰則, "調停要約のための正解コーパスの作成とその分析," 言語処理学会第 17 回年次大会発表論文集 (2011).
- [7] M. Nakano, H. Shibuki, R. Miyazaki, M. Ishioroshi, K. Kaneko, T. Mori, "Construction of Text Summarization Corpus for the Credibility of Information on the Web," *Proc. of the 7th Language Resources and Evaluation Conference* (2010).
- [8] C.-Y. Lin and E. Hovy, "Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics," *Proc. of the Human Language Technology Conference 2003* (2003).
- [9] S. Akamine, D. Kawahara, Y. Kato, T. Nakagawa, Y. I. Leon-Suematsu, T. Kawada, K. Inui, S. Kurohashi and Y. Kidawara, "Organizing Information on the Web to Support User Judgments on Information Credibility," *Proc. of the 4th International Universal Communication Symposium* (2010).
- [10] K. Murakami, E. Nichols, J. Mizuno, Y. Watanabe, S. Masuda, H. Goto, M. Ohki, C. Sao, S. Matsuyoshi, K. Inui and Y. Matsumoto, "Statement Map: Reducing Web Information Credibility Noise through Opinion Classification," *Proc. of the Fourth Workshop on Analytics for Noisy Unstructured Text Data* (2010).
- [11] 藤井 敦, "OpinionReader: 意思決定支援を目的とした主観情報の集約・可視化システム," 電子情報通信学会論文誌, Vol. J91-D, No.2 (2008).
- [12] A. Finn, N. Kushmerick and B. Smyth, "Fact or fiction: Content classification for digital libraries," *Proc. of Joint DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries* (2001).
- [13] 河合 剛巨, 岡嶋 穰, 中澤 聡, "Web 文書の時系列分析に基づく意見変化イベントの抽出," 言語処理学会第 17 回年次大会発表論文集 (2011).
- [14] R. Miyazaki, R. Momose, H. Shibuki and T. Mori, "Using Web Page Layout for Extraction of Sender Names," *Proc. of the 3rd International Universal Communication Symposium* (2009).
- [15] 加藤 義清, 河原 大輔, 乾 健太郎, 黒橋 禎夫, 柴田 知秀, "Web ページの情報発信者の同定," 人工知能学会論文誌, Vol. 25, No. 1 (2010).
- [16] 橋本 力, 奥村 学, 島津 明, "複数記事要約のためのサマリパッセージの抽出," 言語処理学会第 7 回年次大会発表論文集 (2001).
- [17] 石下 円香, 佐藤 充, 森 辰則, "Web 文書を対象とした質問の型に依らない質問応答手法," 人工知能学会論文誌, Vol. 24, No. 4 (2009).
- [18] K. Shinzato, T. Shibata, D. Kawahara, C. Hashimoto and S. Kurohashi, "TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology," *Proc. of the Third International Joint Conference on Natural Language Processing* (2008).
- [19] 岡嶋 穰, 河合 剛巨, 中澤 聡, 村上 浩司, 松吉 俊, 水野 淳太, エリック・ニコルズ, 渡邊 陽太郎, 乾 健太郎, 渋木 英潔, 中野 正寛, 宮崎 林太郎, 石下 円香, 森 辰則, "Web 文書の時間・論理関係分析に基づく情報信頼性判断支援システムの開発と実証実験," 言語処理学会第 17 回年次大会発表論文集 (2011).