

調停要約生成の精度向上に向けた手がかり表現の利用と重要度計算手法の検討

永井 隆広* 石下 円香† 中野 正寛† 松本 拓也† 渋谷 英潔† 森 辰則†

*横浜国立大学 大学院 環境情報学院 †横浜国立大学 大学院 環境情報研究院 ‡横浜国立大学 工学部
E-mail: {nagadon,ishioroshi,nakano,m_takuya,shib,mori}@forest.eis.ynu.ac.jp

1 はじめに

Web で多くの人が情報発信を行う現在、Web 上の情報の信憑性を利用者が判断することを支援する技術に対する要求が高まっている。我々は、情報信憑性判断を支援するための技術として、調停要約を提案している [1, 2]。調停要約とは、一見対立しているように見えるが実際は両立可能な関係にある二言明¹の根拠や成立条件、両者が共存できる状況に関する記述が含まれた要約を利用者へ提示することで、信憑性判断の支援を目指す要約である。我々は「ディーゼル車は環境に良い」といった、利用者が信憑性を判断したい言明(以降「着目言明」とする)を Web 検索して得られた文書群を要約対象として、それらから、上記の根拠、成立条件、共存状況などの記述がまとめて書かれたパッセージを抜粋することにより調停要約の生成を行うシステム HERMeS²を開発した [1]。HERMeS は、上記記述の各々を近似的に見つけるために、クエリを変えて Web 文書を再検索し、それに基づき特徴語の重みづけを行い、パッセージの重要度を計算する。

本稿では、HERMeS の重要度計算において、用いる各パラメタが、精度に与える影響の大きさを確認するために、パラメタの値と精度との関係を調査した結果について述べる。また調停要約生成精度向上に向けて、HERMeS の重要度計算に逆接表現や条件表現などの手がかり表現を組み込んだ手法について検討を行う。

2 関連研究

情報信憑性判断支援に関する研究は調停要約以外にもいくつか行われている。赤峯ら [3] は Web 情報の全体像を情報内容、情報発信者、情報外観などの観点から集約して提示することで、利用者が Web 情報の信頼性を検証するためのシステム「WISDOM」の開発を行っている。Murakami et al. [4] は言明間の対立関係や根拠関係などの解析を行い、言明間の関係を俯瞰することができる「言論マップ」を開発している。これらの研究と調停要約は、言明間の関係の理解を支援するという点は同じであるが、対立の読み解き方を要約を用いて利用者へ提示する点で異なる。

また、阪野ら [5] は日本語判決文特有の手掛かり表現を利用した判決文の自動要約手法を提案した。本稿でも要約に手掛かり表現を利用するが、本稿で用いる手掛かり表現は、調停要約抽出に特化した逆接や条件の表現である点で阪野らと異なる。

3 調停要約

調停要約は、着目言明に関する、Web 上の情報信憑性を利用者が判断するための手がかりとなるような要約である。通常の文書要約では、要約対象の文書が入力と

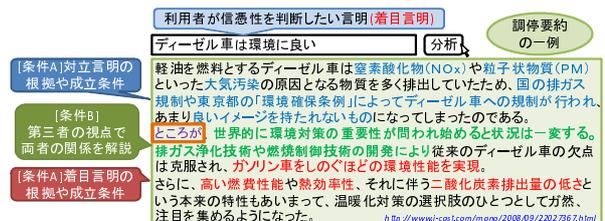


図 1: 調停要約の一例

して与えられるが、調停要約では着目言明が入力として与えられ、着目言明を用いた Web 検索により得られた文書集合を要約対象としている。これは、Web 検索エンジンにクエリを与えるように、我々のシステムに利用者が着目言明を与えることを想定し、それに応じて判断の手がかりとなる Web 文書を収集・要約するシステムを研究開発したいという我々の考えによるものである。我々は理想的な調停要約は、以下の 2 条件の両者を満たすような文書あるいは文書群であると考えている。

- 条件 A: 対立している二言明の両者について、根拠や成立条件が示されていること
- 条件 B: 第三者の視点で両者が共存できることが書かれていること

「ディーゼル車は環境に良い」という言明の調停要約の一例を図 1 に示す。根拠や成立条件、第三者の視点などは複数の観点が存在すると考えられることから、調停要約は理想的なものを 1 つ提示すれば良いわけではなく、複数の調停要約を提示し、複数の観点を網羅する必要があると考えられる。理想的な調停要約の条件を満たすようなパッセージ群を、検索された Web 文書から抽出・提示することが出来れば、利用者はなぜ対立が生じたかを理解しやすくなり、着目言明の信憑性の判断が容易になると考えられる。

4 調停要約生成手法

Web 上には、ある対立関係について解説しているような、調停要約としての条件を満たす記述が存在する場合がある。入力された着目言明に基づいて検索された Web 文書を要約対象とし、上記のような記述をパッセージ単位で抽出することによって生成される調停要約のことを直接調停要約と呼ぶ。一般的な複数文書要約と異なり、直接調停要約は、一文書からまとめた部分を抽出して作られる要約なので、文の結束性や意味的整合性が担保される。

3章で述べた理想的な調停要約の 2 つの条件のうち、条件 A に関する記述は、対立する言明に現れる語など

²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

²Helpful Explanation Representing Mediator Summary のイニシャルを取り、命名した。

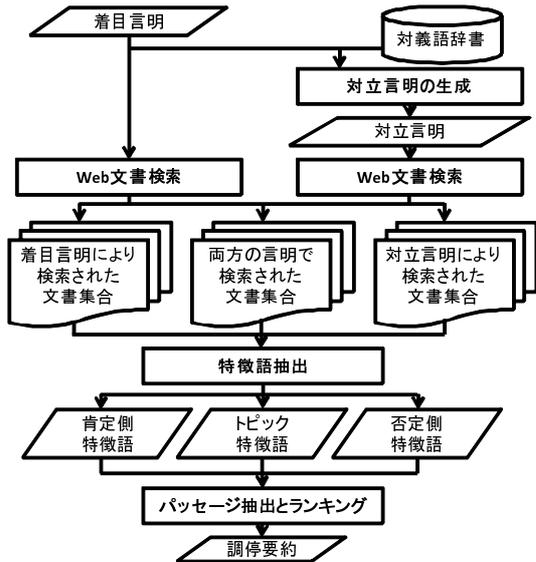


図 2: HERMeS の処理の流れ

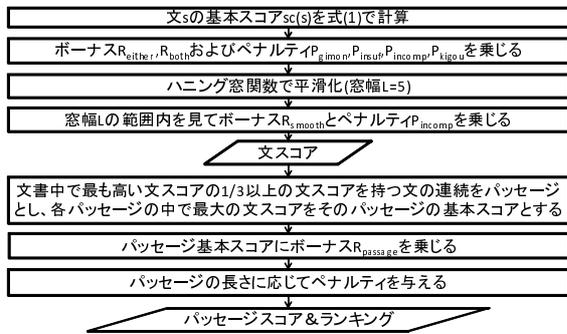


図 3: パッセージ抽出とランキングの流れ

を手掛かりとして抽出できると考えられる．一方，条件 B に関する記述は，条件や根拠の種類などによって記述のされ方が様々であり，これを直接の抽出対象とすることは難しいと考えられる．我々は条件 A に関する記述の周囲に，条件 B に関する記述があるのではないかと考えた．HERMeS は，上記の考えに基づいて直接調停要約パッセージ群の抽出を目指す．

調停要約生成の全体の流れを図 2 に示す．ここでは，紙面の都合から本稿に関連する部分のみ，簡単に説明する．各処理の詳細については，文献 [1, 2] を参照されたい．肯定側・否定側特徴語の抽出について説明する．検索された全ての文書における単語の頻度に基づいてランキングを行い，上位 C_{rank} 位に含まれる単語を肯定側・否定側特徴語の候補とする．次に着目言明と対立言明により検索された文書集合から，各単語に肯定側スコアと否定側スコアを付け，それぞれランキングする．特徴語候補のうち，肯定側順位が否定側順位よりも閾値 C_{dif} を超えて大きい場合には肯定側特徴語とし，否定側順位が肯定側順位よりも C_{dif} を超えて大きい場合には否定側特徴語として抽出する．

パッセージ抽出とランキングは図 3 の流れで行う．まず各文に式 (1) によりスコアを付ける．

$$sc(s) = \frac{\sum_{w \in KW} appear(w, s)}{|KW|} \quad (1)$$

KW は 3 種類の特徴語の和集合， $appear(w, s)$ 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

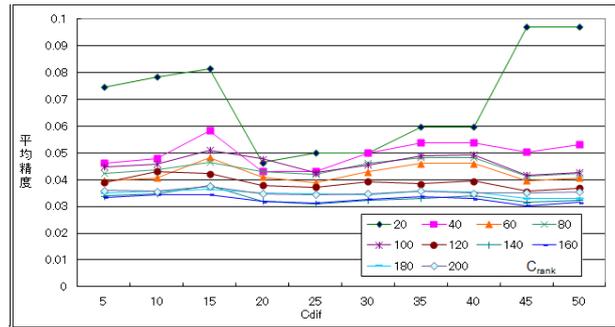


図 4: パラメタ C を変化したときの平均精度 (原発)

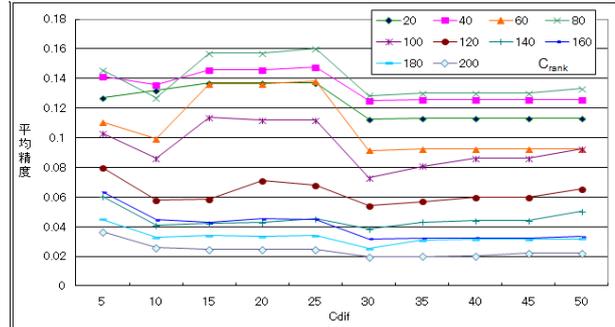


図 5: パラメタ C を変化したときの平均精度 (嘘)

返す関数である．重要度計算で用いるパラメタには図 3 中の各ボーナス R とペナルティ P があり，それぞれ表 1 の役割を持っている．各パラメタは文 s それぞれの条件を満たすような文である場合に乘じられる．各ボーナス R は 1 以上の定数で，各ペナルティ P は 1 以下の定数である．パッセージスコアを基にランキングを行い，調停要約パッセージ群として出力する．

5 重要度計算における各パラメタ値の調査

我々は，4 章で述べた重要度計算において，精度が表 1 の各パラメタにどの程度依存するかを調査した．調査には，調停要約コーパス [2] を利用した．コーパスには，幾つかの着目言明に関する文書が収録されており，各文書にはどこが調停要約パッセージであるかの情報が付与されている．正解となる調停要約パッセージは人手による抜粋により作られ，不正解パッセージは正解パッセージを含まない文書を数文ずつ分割することで作られている．

使用した着目言明，コーパスに含まれるパッセージの総数，その中の正解パッセージの数は，表 2 のとおりである．はじめに我々は，各パラメタの最適値を求めたいと考え，各パラメタを変化させたときの精度の変化について予備調査を行った．予備調査の結果，精度変化の傾向が着目言明によって異なっており，最適値を求めることが困難であることが明らかになった．そこで，パラメタ C ， P ， R が精度に与える有効性の調査を行うこととした．まず， P に関するパラメタは 0， R に関するパラメタは 2 で統一し，特徴語抽出に関するパラメタ C_{rank} と C_{dif} を変えたときのシステムの精度を調査した．評価尺度として式 (2) の平均精度 (AP) を用いた．

$$AP = \frac{1}{R} \sum_r r I(r) P(r) \quad (2)$$

R は正解数， $I(r)$ は第 r 位での正解かどうかのフラグ， $P(r)$ は第 r 位での精度である．結果を図 4，図 5 に

表 1: 調査したパラメタの一覧と本調査に基づく各パラメタの設定値

パラメタ名	パラメタの役割	値
C_{rank}	特徴語の候補とする語の順位の下限	80
C_{dif}	特徴語として採用する語が持つべき順位の違いの閾値	15
P_{incomp}	省略された文(「・・・」で終わる文)のスコアに乘じるペナルティ値 (≤ 1)	0
P_{kiyou}	記号ばかりの文のスコアに乘じるペナルティ値 (≤ 1)	0
P_{qimon}	「？」で文が終了している文のスコアに乘じるペナルティ値 (≤ 1)	1
P_{insuf}	不十分な文(動詞, 名詞が一定数以下の文)のスコアに乘じるペナルティ値 (≤ 1)	0
R_{either}	トピック特徴語に加え肯定 or 否定特徴語に言及している文のスコアに乘じるボーナス値	2
R_{both}	3 種の特徴語に言及している文のスコアに乘じるボーナス値	2
R_{smooth}	当該文を中心とした窓幅 L の範囲内に含まれる文列が 3 種の特徴語に言及している場合に当該文のスコアに乘じるボーナス値	2
$R_{passage}$	3 種の特徴語に言及しているパッセージのスコアに乘じるボーナス値	2

表 2: 利用した着目言明とパッセージ数

着目言明	総数	正解
原発は地震でも安全である	8,092	57
嘘をつくのは悪いことである	7,813	36

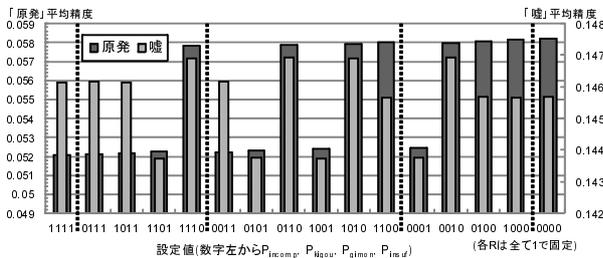


図 6: パラメタ P を変化させたときの平均精度

示す．結果より， C_{rank} はあまり大きすぎると精度が下がってしまう傾向にあり， C_{dif} の傾向は着目言明によりかなり異なることが分かった．

次に， C_{rank} と C_{dif} を比較的良好な平均精度だった値 ($C_{rank} = 40$, $C_{dif} = 15$) に固定し， P に関するパラメタを 0 または 1 に， R に関するパラメタを 1 または 2 に変化させた時の平均精度を調べた．結果を図 6，図 7 に示す．結果より， P に関するパラメタは平均精度に対して極めて小さい影響しか与えないことが分かった．影響が小さかったのは，この実験においてパッセージがあらかじめ切り分けられていたためと考えられる．また， R に関するパラメタは R_{either} ， R_{both} ， R_{smooth} ， $R_{passage}$ のいずれも 2 にした時，つまり特徴語が含まれているときのボーナスを入れたときが最も平均精度が高いことが分かった． R に関する各パラメタを 2 以外に変化させる調査も行ったが，パラメタを 2 にした時と精度は大きく変わらなかった．このことから R に関する各パラメタは精度向上へ寄与するが，その値による精度への影響は小さいことが分かった．これらの結果から，表 1 の各パラメタの値を今後の実験では用いることとした．

6 手掛かり表現の利用

我々は，文献 [2] において，HERMeS の出力パッセージに対して「しかし」や「けれど」のような逆接表現が含まれているパッセージを残すようなフィルタリング処理を行い，出力精度が向上することを確かめた．しかし，フィルタリング処理を行うことにより，本来正解にすべきパッセージまで除外してしまうという問題点

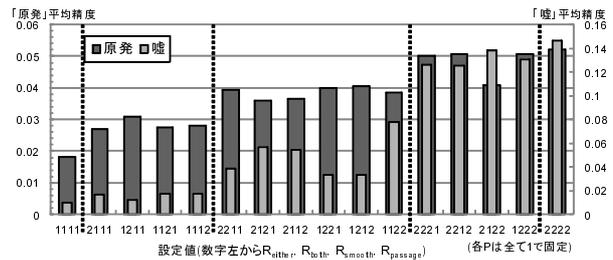


図 7: パラメタ R を変化させたときの平均精度

があった．また，調停要約の正解例を調査すると，「但し」「A の場合 B」のような条件を表す表層表現も多く出現したことから，これらも手掛かり表現として使うべきではないかと考えた．そこで我々は，4 章の，文やパッセージの重要度計算において逆接表現や条件表現を組み込んだ手法を提案する．今回手掛かり表現として利用した語句を表 3 に示す．表中の太字の表現は文献 [2] で使用した表現 (以下，旧表現とする) であり，太字になっていないものは今回追加した手掛かり表現 (以下，新表現とする) である³．

手掛かり表現の拡充と重要度計算における手掛かり表現の利用が有効であるか確かめるための評価実験を行った．以下の 7 つの手法について実験を行う．

手法 A: 4 章の処理による手法

手法 B: 手法 A に対して，旧表現を用いてフィルタリングを行う手法 (文献 [2] での手法に相当)

手法 C: 手法 A に対して，新表現を用いてフィルタリングを行う手法

手法 D: 式 (1) を式 (3) に置き換え， $R_{passage}$ の条件に旧表現を含むことを加えた手法

手法 E: 式 (1) を式 (4) に置き換え， $R_{passage}$ の条件に旧表現を含むことを加えた手法

手法 F: 式 (1) を式 (3) に置き換え， $R_{passage}$ の条件に新表現を含むことを加えた手法

手法 G: 式 (1) を式 (4) に置き換え， $R_{passage}$ の条件に新表現を含むことを加えた手法

$$sc(s) = \frac{\sum_{w \in KW} appear(w, s) + Q_{flag}(s)}{|KW|} \quad (3)$$

$$sc(s) = \frac{\sum_{w \in KW} appear(w, s)}{|KW|} (Q_{flag}(s) + 1) \quad (4)$$

$Q_{flag}(s)$ は手掛かり表現が文 s 中に 1 つでもあれば 1，そうでなければ 0 を返す関数である．

³この表現一覧は，本稿での評価に用いていない言明に関する調停要約や，文献 [6] を参考に作成した．

表 3: 手掛かり表現一覧

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, が, だけど, ただ, それなのに, にもかかわらず, 但し, ただし
接続助詞	けれど, けれども, (仮定形) + ば
格助詞	に対し, に対して
名詞	反面, 他方, 限り, 場合
複合語	逆に, 反対に, よりも, の方が, のほうが, に比べる, と比べる, と比較する, だけだない, ではない, にもかかわらず

表 4 の 10 個の着目言明⁴について, 各手法を用いて調停要約を生成し, 第 1, 第 3, 第 4 著者の 3 名で評価を行った. 1 着目言明あたりの出力パッセージ数は 10 とした. つまり 1 手法あたり 100 パッセージが評価対象となる. 各手法の適合パッセージ数と累積平均 MRR⁵を表 5 に示す.

表 4: 評価実験に使用した着目言明

赤ちゃんポストは必要だ
就活にスマートフォンは必要である
農業は体に悪い
水をたくさん飲むと健康に良い
北海道新幹線は必要である
地球温暖化によって海面が上昇する
死刑制度は廃止すべきだ
TPP に参加したほうが良い
節電で経済が衰退する
実名登録は危険だ

表 5: 適合パッセージ数と累積平均 MRR

手法名	適合数	累積平均 MRR
手法 A	8	0.10
手法 B	10	0.31
手法 C	9	0.19
手法 D	10	0.28
手法 E	13	0.40
手法 F	12	0.26
手法 G	9	0.30

表 5 より, 手掛かり表現を考慮しない手法 A に比べ手法 B ~ G は適合数, 累積平均 MRR 共に上昇している. このことから, 手掛かり表現が有効に機能していることがわかる. 手法 B と手法 C を比べると適合数, 累積平均 MRR とともに手法 B の方が高い. これは手掛かり表現を増やしたことで, 旧表現では取り除くことが出来た不要なパッセージを残してしまうようになったためと考えられる. 適合数, 累積平均 MRR とともに最も良い結果となったのは, 手法 E であった. このことから, 従来のフィルタリング処理から, 重要度計算に手掛かり表現を組み込んだ処理に変更したことにより, 不採用にしていた正解パッセージを残すことが出来るようになったと考えられる. 一方で, 手掛かり表現の拡充は, 我々の予想に反して逆に精度を下げってしまう結果になった. これは, 今回試した手法では, いずれも逆接と条件の表現を同一視して扱っているためではないかと考えられる. そのため, 今後は逆接と条件の表現を区別し, 対義語などの出現位置などを考慮することが考えられる.

また, 今回, 調停要約がありそうな着目言明を選んだにもかかわらず, 正解が 100 パッセージ中 10 個程度と極めて低くなってしまい, 最も精度が良かった手法 E でも 10 着目言明のうち 3 着目言明では, 正解が 1 つも見つからなかった. 例えば, 「北海道新幹線」の言明で, 九州新幹線が必要かどうかを述べているパッセージ

⁴事前の調査で調停要約が存在しそうなものを選択した.

⁵正解が見つかった順位の逆数を加算して, その平均値を全着目言明について計算した結果で, 値が高いほど上位に正解を出力出来ることになる [1].

が出力されたり, 「農業」の言明で有機栽培の野菜の安全性について述べているパッセージが出力されているなど, 着目言明と関連はしているが, 利用者が判断したい内容と異なる内容を述べているようなパッセージが多く見受けられた. このような焦点がずれてしまう問題を解決するために, 我々は, 現在, ユーザーが焦点を選択するような調停要約生成システムを検討中である. 加えて, より着目言明に則した調停要約生成ができるように, 特徴語抽出手法を改良することも考える必要があると思われる.

7 おわりに

本稿では, 調停要約生成における重要度計算に用いる各パラメタ値の検討と, 逆接と条件を表す手掛かり表現の利用に関する検討を行った. パラメタ値に関しては, 着目言明による影響が大きく最適値を求めることが困難であったため, 各パラメタの精度に与える影響の傾向を確認し, 比較的妥当と思われるパラメタ値を利用することとした. 手掛かり表現の利用に関しては, 従来手法のフィルタリングから本稿で提案した重要度計算における利用に変更することで精度向上を確認した. しかし, 条件表現などの手掛かり表現の拡充には, 改善効果が見られなかった. 今後の課題として, 対義語の出現位置などを考慮した重要度計算手法のさらなる改善と特徴語抽出の改善を行いたいと考えている.

謝辞

本研究の一部は科学研究費補助金 (No.22500124), ならびに横浜国立大学大学院環境情報研究院平成 23 年度共同研究推進プログラムの助成を受けたものである.

参考文献

- [1] 中野正寛, 渋木英潔, 宮崎林太郎, 石下円香, 金子浩一, 永井隆広, 森辰則. 情報信憑性判断支援のための直接調停要約生成手法. 電子情報通信学会論文誌, Vol. J94-D, No. 11, pp. 1919-1930, 2011.
- [2] 渋木英潔, 中野正寛, 石下円香, 永井隆広, 森辰則. 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. 第 10 回情報科学技術フォーラム (FIT 2011) 講演論文集, pp. RE-003, 2011.
- [3] 赤峯亨, 宮森恒, 加藤義清, 中川哲治, 乾健太郎, 黒橋禎夫, 木依豊. Web 情報の信頼性検証のための情報分析システム WISDOM. 言語処理学会第 14 回年次大会発表論文集, pp. 721-724, 3 月 2008.
- [4] Koji Murakami, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments. In *the 3rd Workshop on Information Credibility on the Web (WICOW2009)*, pp. 43-50, 2009.
- [5] 阪野慎司, 松原茂樹, 吉川正俊. 手がかり表現に基づく判決文の自動要約. 言語処理学会第 11 回年次大会発表論文集, pp. 193-196, 2005.
- [6] 石黒圭. 文章は接続詞で決まる. 光文社, 9 2008.