

引用意見とその発言者を利用した意見情報の提示

佐藤 沙紀[†] 関 洋平^{††}

[†] 筑波大学情報学群知識情報・図書館学類 〒305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学図書館情報メディア系 〒305-8550 茨城県つくば市春日 1-2

E-mail: [†]ts1011556@u.tsukuba.ac.jp, ^{††}yohei@slis.tsukuba.ac.jp

あらまし 論説型レポートの執筆や、政策判断のための世論調査を行うために、Web 上の意見情報を有効に活用することが望まれている。本研究では、ユーザの情報要求に適合した意見情報を収集・整理することを目的として、発言者の立場に着目した意見情報の提示手法を提案する。まず、ニュースサイトを中心に Web 記事を収集し、記事中の引用部分“「…」”に着目し、その中に含まれる意見を抽出する。次に、引用意見の周囲に現れる人名や組織名を、発言者として抽出することで、立場に基づいた意見情報の分類を試みた。いくつかのクエリについて調査した結果、特に複数の立場に分かれて論争が行われている話題について、提案手法が有効であることを明らかにした。

キーワード 意見抽出, 発言者, 引用意見

1. 研究の目的

現在、インターネットを通じて多くのメディアから情報を獲得することが可能になっている。このような状況下において、自身の望む情報を取得するためには、ニュースサイト、ブログといった様々なメディアを検索する必要がある。しかし、情報収集を行う際、情報要求に応じて収集すべき情報は異なる。Web 上のメディアから、情報を集めて利用する状況の例として、レポートの作成が挙げられる。特に、議論性のあるレポートを作成する際は、様々な情報を集める必要があると考えられる。例えば、「原子力発電を行うことに対して賛成か、反対か」といったテーマでレポートを作成する際、原子力発電についての詳細な事実情報を集めるだけでなく、原子力発電を行うことへの意見も必要となってくる。このように自身の望む情報を取得する際には、事実情報と意見情報の 2 種類に分けて情報を集める必要がある。ここでいう事実情報とは、客観的に説明できる事象を示し、意見情報とはある特定の個人（組織）の考え、主張を示す。議論性のあるテーマについてレポートを作成する際、事実情報は Google^(注1) といったサーチエンジンを使用することで比較的容易に収集することができる。しかし、自身の主張と関連のある意見情報を検索エンジンを用いて適切に収集することは難しい。また、意見情報を収集し、ランキング表示を行っても、自身の主張と関連のある意見がまとまって提示されることはない。そのため、ランキング表示を行うのではなく、自身の主張との関連の有無で意見情報を分類し、提示する必要がある。本研究では、意見情報の発言者や、発言者が所属する組織を立場とし、組織ごとあるいは発言者ごとに意見情報をまとめて提示する手法について検討する。

現在、Web から意見を収集し、提示しているシステムとして、WISDOM^(注2) や Yahoo! ニュース^(注3) が実用化された例と

して挙げられる。WISDOM は Web 情報を様々な観点から分析・集約し、提示するシステムであり、これを使用することで肯定意見か否定意見かを分析することができる。肯定意見か否定意見かを分析することで意見の分布図を提示したり、肯定意見一覧、否定意見一覧として提示したり、どのようなサイトに意見情報が多く含まれているのかを提示したりしている。これにより、肯定意見か否定意見かという観点ごとに意見を見ることが出来る。一方、Yahoo! ニュースでは、通常のニュース記事の下部にコメントを表示することができる。コメント機能は、Yahoo! アカウントを使用するもののみならず、Facebook コメント、Twitter のつぶやきも表示される。これらのコメントを表示することで、記事を読んだ個人の意見が一覧となって表示され、記事内容のみならず、他者の意見も同時に知ることができるという利点がある。これらの実用化されているシステムで実際に意見情報を利用していることから、意見情報を抽出し、提示することには利用価値があると考えられる。しかしながら、レポートを作成する際には、肯定か否定かで意見を分類することが必ずしも有効とは限らない。肯定意見として述べられている事柄を根拠として、新たな反論意見を考え出すこともあるためである。本研究では、ニュース記事を主な対象として、あるクエリに適合する引用意見を抽出し、発言者ごとに意見を分類する方法の有効性について検証する。

2. 関連研究

2.1 意見情報の抽出に関する研究

意見情報の抽出は、主にネット上のテキストを題材として行われている。小林ら [4] によると、意見抽出の中でも、文書分類に基づくアプローチとして、テキストや段落 (あるいは文) の単位で肯定意見か否定意見かといった意見クラスに分類する課題を設定し、機械学習に基づく様々な手法の実験結果が報告されてきたとある。本研究もそのような文書分類に基づくアプローチを取るものの 1 つである。小林らは、ネット上に分散する個人の意見を検索・分類・要約する意見マイニングを目的と

(注1) : <https://www.google.co.jp/>

(注2) : <http://wisdom-nict.jp/>

(注3) : <http://news.yahoo.co.jp/>

して研究を進めている。意見情報の抽出の課題設定に関して、(評価対象, 属性, 評価値) からなる従来の構造化の枠組みを拡張し, 新しい使用に基づいた意見タグ付きコーパスを使用することで, 意見抽出の範囲設定や構造について考察を行っている。この考察の中で, 抽出すべき意見として, 「商品やサービス, 人物, 組織, 政策など, 特定のクラスの実体に対して言語的に表明される意見」とある。本研究では抽出する意見として, 特に人物, 組織, 政策に関連する意見情報に焦点を当てる。

次に, 意見情報が多く含まれる媒体である新聞のテキストを対象としている研究として, Balahur ら [1] [2] [3] が挙げられる。Balahur らはニュース記事の引用意見文 (“…”) に着目し “target” (あるテーマ) が規範に沿っているか反しているかという分類をおこない, 良いニュース, 悪いニュースを判断するとともに, 肯定的な感情と否定的な感情を識別し, ニュースコンテンツの良し悪しを決定する手法を考案した。新聞記事からの意見情報の抽出に関しては, 引用部分にテキストの他の部分よりも意見情報が多く出現しているため, 引用意見の抽出がそのまま意見情報の抽出になると述べられている。本研究でも, ニュース記事を対象として, その周囲に発言者などの立場を表す言葉が現れやすい引用意見を抽出し, 立場ごとに分類することについて検証する。

また, 意見情報を肯定意見か否定意見かという分類を行い, 提示を行っている研究として WISDOM^(注4) が挙げられる。これは, 述語項構造に基づき, あるトピックに関する主要・対立表現を俯瞰的に提示するシステムとなっている。WISDOM に関する研究として, 赤峯ら [5] や宮森 [6] ら, 河原ら [9] が挙げられる。Web 上の情報の信頼性を情報内容, 情報発信者, 情報外観, 社会的評価といった 4 つの基準でとらえることを提案しており, これらを述語項構造を単位とする自然言語処理によって論理的に分析・組織化することが目指されている。

文書内の対立構造に着目した研究として, 中野ら [8] や永井ら [11] は着目言明に関して, 対立関係にあるように見える記述(疑似対立)を読み解く手掛かりを簡潔に提示する調停要約を提案している。中野らは, ある主観的な意見, 評価, 疑問の表明や客観的事実を含めたテキスト情報を言明とし, 一つの着目言明に対する対立関係の読み解きを試みた。ある条件や状況下において互いの内容が両立できる関係を疑似対立と呼び, その対立を読み解くための手がかりを簡潔に示す調停要約を提案している。これらの研究から, Web 上の意見情報を集約し, 提示することについての需要があるということが出来る。本研究では, 引用意見の発言者ごとに意見を集約することが, こうした需要にどの程度応えるものであり, またどのような課題があるかを明らかにすることを目的とする。

2.2 立場の抽出に関する研究

意見情報を抽出し, 分類を試みる際, 現在は肯定意見か否定意見かという 2 つでの分類が行われているが, 本研究では別の分類を試みる。具体的には, 意見情報の発言者に着目し, 発言者の個人名, 組織名を立場とし, 立場ごとに意見情報を分類す

ることを試みる。このように意見情報の発言者に着目されている関連研究としては, 以下のものが挙げられる。吉岡ら [7] は日中韓を中心とした複数国の新聞記事をサイトから収集し, 注目をあびている単語リストや, そのような単語を含む記事数の増減, 関連語を表示し従来とは異なる観点からの分析を可能にする手法を提案した。この研究は, 意見分析に着目しているわけではないが, 同じトピックについて, ニュース配信社ごと, あるいは異なる国において選択されている話題の差異を提示しており, ある一つの配信社だけからは得ることのできない知識を, ユーザは新たに獲得することができる。このように, 同じ話題について複数の立場を比較することは, 知識獲得の面から有用である。本研究では, 同じ話題についての意見を発言者という立場ごとに整理することの有用性について検証する。

また, 関ら [10] は, 意見を表明する主体を意見保有者とし, 判別することで, 見方を区分した意見集約が行えるとして, 意見文における著者と第三者の観点の違いを, 書き方のスタイルに着目して区別する手法を提案している。同時に, あらゆる意見を対象とした意見保有者の抽出は難しいタスクであることも明らかにしている。本研究では, 対象を引用意見に絞り込み, 発言者の抽出並びに発言者を単位とした意見情報の集約がどの程度有効であるか, 検証を進める。

3. 提案手法の概要

本研究では, Web 上のニュースサイトから意見情報を収集し, ユーザの情報要求に応じて適切に整理した形で提示することを試みる。本研究で提供するシステムのユーザは, 自身の主張を述べる際に他者の意見を利用する状況, たとえば論説型レポートの執筆を想定している。しかし, クエリを入力して適合する意見情報を集約する場合, テキスト中のどの部分に意見が含まれるか漠然としている場合がある。また, 意見情報を集約し, 提示する際にユーザが求める情報を素早く見つけるために, 見つけやすい形に整理する必要がある。ここでは, たとえば政策が関係する意見の場合には, 政党ごとの主張なのか, 個人の主張なのかを判断することで意見情報を見やすくできると考える。そこで, 人名や組織名などを立場として判別し, その立場に基づいて意見情報を整理することを提案する。

意見を提示している人名や組織名をその意見の立場とするならば, テキスト中に現れる意見情報の発言者を, 意見情報とあわせて抽出することにより, その意見の立場を判別することができる。通常のニュース記事の中から意見情報を発見し, 立場を抽出しようと試みると, どの部分が意見情報となるのか, 発言者が誰なのかを見分けることは難しい。そこで “「…」” で囲われた部分を引用部分とし, その中に意見が含まれる場合に, その情報を抽出することを試みる。引用意見を示す “「…」” の前後には, その意見を提示している発言者が表れやすいため, 立場の抽出にも適していると考えられる。

本研究では, 以上の議論に基づき, Web サイト上の意見を収集し, 立場ごとに整理する。まず, Web 上のニュースサイトを

(注4) : <http://wisdom-nict.jp/>

起点とし、NICT Web クローラ^(注5)を使用してニュース記事を収集した。次に、収集したデータをもとに、引用部分に着目し、意見情報を抽出した。最後に、人名、組織名、政党名を立場として判別し、立場に基づいて意見情報を分類・提示する。これにより、複数の立場ごとに意見を抽出し、比較分析することが可能となる。

3.1 抽出に使用するデータに関して

本研究で使用するデータについて説明する。本研究では、Web上のニュースサイトを対象として扱うため、ニュース記事を収集し、テキストを抽出してデータを作成する必要がある。そのため、ALAGINが提供するNICT Web クローラを使用し、朝日新聞、毎日新聞、読売新聞の3サイトを起点としWeb上の記事を収集した。データ収集日は2013年11月15日～16日、11月27日～12月16日までの日間、1日2回、8時と20時に収集を行った。収集した記事から、htmlタグ、タブ、データ取得時につけられたタグを削除したテキストを使用する。

3.2 適合意見の抽出

収集したテキストを対象として、ユーザの情報要求に応じ、適合意見を抽出する。まず、クエリを含む引用意見の抽出を行う。次に、クエリを拡張し、関連語を含む引用意見の抽出を行う。クエリの拡張には、kizasiのAPI「kizAPI(きざび)」^(注6)を用いる。

3.3 引用意見の抽出

本研究では、立場ごとに意見情報を整理するために、引用意見を抽出する。具体的には、まず、引用部分(「…」で囲われた部分)を抽出し、その中に含まれる要素が意見であるか否かを判別する。引用部分は、ニュース記事の他のテキストより、主観的である可能性が高いため、意見情報として利用できる主観が含まれている可能性が高い。しかし、テキストから「」で囲われた引用部分を抽出し、その一覧から特徴を考察した結果、文章が短く、名詞を含むものが広告である可能性が高いことがわかった。そこで、引用符に囲まれた部分を抽出し、形態素解析器MeCabにかけ、形容詞あるいは、名詞の中でも形容動詞語幹となる形態素を含み、総形態素数がn件(n=10)以上で、かつ最後の形態素が名詞で終わらないものを、意見情報を含む引用意見として抽出することとした。

3.4 立場の判別

本研究では、引用部分に含まれている意見情報を引用意見とし、抽出を試みる。その際、引用意見の周囲には立場が含まれ、この立場で引用意見を分類・提示すると仮定する。立場ごとに意見を比較するためには、何を立場とするのかという問題がある。ここでは、立場となる語は、“安倍晋三”、“民主党”、“朝日新聞社”というように固有表現(人名、組織名、政党名)となる場合が多いことから、引用意見の周囲に出現する固有表現に着目して立場の抽出を行う。この点については、次章で予備実験を行い、詳細に検討を進める。

4. 予備実験：立場の抽出に関する検証

4.1 実験目的

Web上の記事をもとに作成したデータを用い、引用意見の周辺に出現する人名、政党名、組織名といった立場を抽出できるかを検証することを目的とする。この時、立場は固有表現と仮定し、抽出を試みる。固有表現の発見には、Cabocha^(注7)の固有表現抽出機能を使用し、その中でも人物を表す<PERSON>、組織を表す<ORGANIZATION>を利用して立場を抽出する。また、単に引用意見の周辺の固有表現を抽出するだけではなく、主格を示す助詞や、“「」からの距離などといった手がかりを利用して立場を抽出することについて検証する。

4.2 実験方法

実験に使用する分析対象として、読売新聞^(注8)、朝日新聞^(注9)、毎日新聞^(注10)のwebニュースサイトの記事を使用した。事前にレポート作成時のテーマを3種設定し、各テーマにおいて検索エンジンGoogle^(注11)のサイト検索に出現する上位10件として表示された記事を収集した。レポート作成時のテーマは、以下の通りである。

- (1) 憲法96条の改正について意見を述べよ。
- (2) 体罰問題に関して意見を述べよ。
- (3) ネット選挙解禁に関して意見を述べよ。

予備実験では、これらのテーマに沿って収集したデータから、引用意見を選定し、その周辺に出現する立場の抽出を試みる。具体的な実験方法は以下の通りである。

(1) 引用意見を含む文を30文抽出し、手作業で立場を判別。

(2) Cabochaの固有表現抽出機能を使用し、データから組織名と人名を判別し、以下の3つの手法により立場を抽出し、評価結果を比較する。

A) 係助詞“は”の直前の語を抽出。

B) “「”の前に人名または組織名があれば抽出。候補が複数存在する場合、

- 人名と組織名が含まれていた場合、人名を抽出。
- そうでない場合、“「”からの距離が近いものを抽出。

C) 係助詞“は”から“「”までの距離が15字以内の場合手順Aを行い、次に手順Bを行う、15字以上の場合手順Bを行う。

4.3 評価基準

実験方法に従って抽出を行った結果を、評価するにあたり、3段階で評価基準を設ける。評価基準は以下のとおりである。

- (1) ○ (1点) …人手で判別した立場と同じ立場を抽出できている。
- (2) △ (0.5点) …人手で判別した立場の、一部を抽出。
- (3) × (0点) …人手で判別した立場と、完全に異なる語

(注7) : <https://code.google.com/p/cabocha/>

(注8) : <http://www.yomiuri.co.jp/>

(注9) : <http://www.asahi.com/>

(注10) : <http://mainichi.jp/>

(注11) : <https://www.google.co.jp/>

(注5) : <https://alaginrc.nict.go.jp/resources/nictmaster/software/crawler-info/crawleroutline.html>

(注6) : <http://kizasi.jp/tool/kizapi.html>

を抽出。

△と判断したものの例としては、“公明党憲法調査会長の北側一雄副代表は29日のTBS番組で、「(憲法改正の)手続きの緩和については、議論をしっかりとしたい。ただ、過半数はいかなるものか」と述べ、発議要件の緩和に慎重な考えを示した。”という抽出結果に対して、北川のみが抽出されてしまった場合が挙げられる。

4.4 実験結果

予備実験の結果を表1に示す。手法Cが他の手法に比べて、○の数が多いという結果が得られた。本研究では、この結果に基づき、以降の議論では手法Cに基づき立場の抽出を行うこととする。

手法	A:“は”の直前	B:人名優先	C:A+B
精度	0.36	0.20	0.50

5. Webからの引用意見の抽出と立場の判別

5.1 実験目的

Webから収集したデータを対象として、情報要求に適合した引用意見を抽出できているか、立場を判別できているかを調査する。使用するデータは、ALAGINが提供するNICT Webクローラ^(注12)を使用し、読売新聞^(注13)、朝日新聞^(注14)、毎日新聞^(注15)の3サイトを起点として、Web上の記事を収集したものとする。データ収集日は2013年11月15日～16日、11月27日～1月31日までの間、1日2回、8時と20時に収集を行った。

5.2 実験方法

最近、ニュース等で取り上げられ、意見が活発に交わされていると考えられる話題をもとにレポート作成用のテーマを設定し、そのテーマに適合する引用意見を抽出できているかどうかを評価する。また、判別した立場が、引用意見に対して適切か評価する。実験に使用するレポート作成用のテーマと、各レポートテーマに対し、意見情報の収集に使用するクエリは以下の通りである。

表2 レポート作成用のテーマと設定したクエリ

レポート作成用のテーマ	クエリ
北朝鮮の張氏処刑問題に関して見解を述べよ	北朝鮮
消費税の増税に関して意見を述べよ	消費税
秘密保護法の成立に関して意見を述べよ	秘密保護法
原発問題に関して意見を述べよ	原発
猪瀬都知事の辞任に関して意見を述べよ	猪瀬

次に、これらの各クエリに対し、関連語を取得する。クエリ

の拡張のために、kizasiのAPI「kizAPI(きざっぴ)」^(注16)を用い、関連語の一覧を取得する。取得した関連語の中には、“日本”、“容疑者”といったように、汎用性が高く、適合する引用意見を取ることが難しいと考えられる語も含まれている。そのため、手作業で関連語の絞り込みを行う。汎用性が高い、または最初に設定したクエリを使用して抽出した引用意見よりも、他の話題に対する引用意見とのかかわりが深いと考えられる語を除き、特にクエリと関連があると思われる語のみを使用した。

これらのクエリとその関連語を使用し、引用意見の抽出と立場の判別を行った。クエリまたは関連語を使用した具体的な実験方法は以下の通りである。

(1) Webから収集したデータを使用し、引用意見中を含む同一文中に、クエリまたは関連語を含む文を抽出する。引用意見は、3.3節の議論に基づき、MeCab^(注17)を用いた結果の引用部分の総形態素数が10以上で、かつ最後の形態素が名詞以外のもの、かつ形容詞または名詞-形容動詞語幹となる形態素を含むものを、抽出する。

(2) Cabochaの固有表現抽出機能を使用し、引用意見の周囲に出現する組織名または人名を、4.2節の手法Cに基づき、立場として抽出する。

(3) 5つのクエリに対し、それぞれ抽出した20件の結果を評価し、比較する。

5.3 評価基準

実験方法に従って抽出を行った結果を、評価するにあたり、評価基準を設ける。本実験では、抽出した引用意見が、レポート作成用のテーマに適合しているかと、抽出した引用意見に対する立場が適切かという2点についてそれぞれ評価する。そのため、引用意見の抽出に対する評価基準と、立場の判別に対する評価基準をそれぞれ設ける。2つとも3段階で評価基準を設けてある。実験に使用した評価基準は以下の通りである。

引用意見の抽出に関する評価基準

(1) ○ (1点) …レポート作成用のテーマに適合した引用意見が抽出できている。

(2) △ (0.5点) …レポート作成用のテーマに適合しては不在が、引用意見が抽出できている。

(3) × (0点) …引用意見を抽出できていない

立場の判別に関する評価基準

(1) ○ (1点) …抽出した引用意見に対して、適切な立場を判別できている。

(2) △ (0.5点) …抽出した引用意見に対して、立場が判別されているが、適切ではない。

(3) × (0点) …立場が判別されていない。

引用意見の抽出に関する評価が△と判定したものの例としては、レポート作成用のテーマ“消費税の増税に関して意見を述べよ”に対し、“監修の奥村先生は、以下のようにコメントした。「抗凝固療法は、多くの患者さんが長期間、服薬をつづける必要があります。食生活や服薬管理の面などで、治療継続を

(注12) : <https://alaginrc.nict.go.jp/resources/nictmastar/software/crawler-info/crawleroutline.html>

(注13) : <http://www.yomiuri.co.jp/>

(注14) : <http://www.asahi.com/>

(注15) : <http://mainichi.jp/>

(注16) : <http://kizasi.jp/tool/kizapi.html>

(注17) : <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

負担に感じている場合は、より“我慢しない”治療が可能となる新規経口抗凝固薬も選択肢の一つに加え、かかりつけ医に相談してみるのも良いでしょう”と抽出した場合が挙げられる。この抽出結果は、引用意見を抽出できてはいるが、内容が消費税ではなく、治療費の話となっているため、適合していない引用意見と判断する。

また、立場の判別は、適合した引用意見が抽出できていることが前提条件として必要となる。そのため、引用意見が抽出されていると判断された文を使用し、立場の判別に関する評価基準に基づき、評価を行う。

5.4 実験結果

各評価基準に基づき、実験の評価を行った。引用意見の抽出に関する実験結果を、表3に示す。また、立場の判別に関する実験結果を、表4に示す。

表3 引用意見の抽出に関する実験の結果

クエリ	北朝鮮	消費税	秘密保護法	原発	猪瀬
精度	0.50	0.40	0.75	0.82	0.40

表4 立場の判別に関する実験の結果

クエリ	北朝鮮	消費税	秘密保護法	原発	猪瀬
精度	1.00	0.87	0.53	0.25	0.75

クエリ“北朝鮮”と“猪瀬”を使用した場合、結果に抽出した20件ではなく、10件を使用している。これは、レポート作成用のテーマとして設定した話題が最新のものであり、Webをクロールしてデータを収集した段階では十分に意見が獲得できておらず、抽出できた引用部分に限度があったためである。

5.5 考察

5つのクエリに対する引用意見の抽出に関する実験結果から、クエリ“秘密保護法”や“原発”で良い結果が得られることがわかった。抽出できた適合する引用意見の例として、“「原発が絶対に安全かと言われるとそうではない。これ以上、原発を増やしていくのは無理だと思う」”や“「原発への依存度を下げ、世界に先駆けて自然エネルギーを推進しないとイケない」”といった文が挙げられる。引用意見の抽出の精度が良かったクエリを見ると、汎用性の高い関連語が少ないという傾向が見られた。また、これらのクエリを用いての結果に広告があまり含まれていなかったことから、広告をうまく取り除くことができれば、適合する引用意見が多くなり、精度が向上すると考えられる。一方、抽出結果に広告が多く含まれた“消費税”は精度が落ちることを確認した。“消費税”に多く含まれた広告の抽出結果例として“総額表示義務に関する特例の適用を受けるために必要となる誤認防止措置に関する考え方(案)”や“「消費税の転嫁を阻害する行為等に関する消費税転嫁対策特別措置法、独占禁止法及び下請法上の考え方(案)」に対する意見の概要とこれに対する考え方」といった法律の見出しが存在した。これは、名詞的要素で終了しているが形態素としては記号と判別したことが影響しており、今後改善する予定である。

また、適合する引用意見に対して行った、立場の判別に関する

実験の結果は、“消費税”や“北朝鮮”といったクエリで良い結果が得られた。クエリ“消費税”を使用して得た結果の具体例として“浜田宏一内閣官房参与(米エール大名誉教授)は15日午後の講演で「日銀の黒田(東彦)総裁が積極的に消費税を上げると言ったのだから、責任とって金融政策はちゃんとやってもらわなければ困る」と述べた。”といった文が挙げられる。この文で判別できた立場は“浜田宏一”となっており適切に判別できていた。

適合する引用意見の抽出では良い結果が得られた、“秘密保護法”や“原発”で立場が判別できなかった理由を検討したところ、これらのクエリを使用した結果には“「政府は原発建設を進めてきたが、この費用を安全な自然エネルギー開発に使い、原発依存度を下げるべきだ」「政府は、原発は低コストだとしてきたが、高レベル放射性廃棄物を処分するには、膨大な費用と数万年単位の時間がかかる」「震災の困難をチャンスに変えることが日本ならできる」”といったように複数の引用意見が連続して現れる例が存在していた。このように、引用意見の周辺に立場となる語が含まれていない場合、立場となる語が省略されているのか、もっと引用意見と離れた場所に明記されているのかを調べ、立場の判別方法を改善する必要がある。

これらの2つの結果を合わせて考えると、“原発”や“秘密保護法”のように、私たちの生活に大きく関連する問題は特に活発に議論が行われており、引用意見も多く抽出することができると考えられる。しかし、“消費税”のように私たちの生活と大きく関係はあるが、関連語の中に“負担”といった汎用性の高い語や、“税金”といった他の話題と関連の深い語が含まれている場合、適合しない引用意見が多く抽出されてしまう。そのためクエリを拡張することで得られた関連語に汎用性の高い普通名詞が少ないあるいは、他の話題よりもクエリとの関連が深い場合、適合する引用意見を抽出できると考えられる。また、適合する引用意見が抽出できても、立場を判別することができない場合は、立場となる語が引用意見の周辺にあるのではなく、より離れた場所に存在しているか、省略されてしまっている可能性がある。このような場合に、立場をどう判別するのかを検討する必要がある。さらに、引用意見の抽出に関して、見出しは異なるが、記事内に使われている引用意見が同じで、何度も抽出されている場合もある。この場合、複数回抽出される引用意見をより強く主張されている引用意見として提示するなど、他の引用意見との差を提示する際に明記する必要があると考える。

6. 本研究のまとめ

本研究では、論説型レポートのテーマを執筆する際に、複数の立場からの意見がWeb上に多数存在する場合、発言者の立場に着目し意見情報を抽出し、提示する手法を提案し、その有用性について検証を行った。ここでは、ニュース記事に含まれる引用部分の中に、ユーザの情報要求に適合する意見情報があると考え、その意見情報を引用意見として抽出した。また、引用意見の周辺には、その意見の発言者が存在すると考え、その発言者である、個人名や、組織名を立場とし、抽出した引用意見をこの立場に基づいて分類・提示することを提案した。

具体的な検証として、Web からクローリングしたデータを用いて、抽出した引用意見の適合性ならびに、抽出した引用意見に対して、適切な立場が判別できているか評価を行った。まず、事前にいくつかレポート作成用のテーマを準備し、そのテーマごとに実験に使用するクエリを設定し、クエリの拡張を行い関連語を得た。次に、設定したクエリ、またはその関連語を含む引用意見を抽出し、立場の判別を試みた。その結果、人の生活に関連が深く、意見交換が盛んに行われている話題に対して、適合意見が抽出できることを明らかにした。

謝 辞

本研究のデータの収集に当たっては、独立行政法人 情報通信研究機構により提供された NICT Web クローラを利用した。ここに深く感謝の意を示す。

本研究の一部は、科学研究費補助金基盤研究 C (課題番号 24500291)、基盤研究 B (課題番号 25280110)、萌芽研究 (課題番号 25540159)、ならびに筑波大学図書館情報メディア系プロジェクト研究による助成に基づき遂行された。

文 献

- [1] Alexandra Balahur and Ralf Steinberger. Rethinking Sentiment Analysis in the News: from Theory to Practice and back. In *WOMSA '09*, pp. 1–12, 2009.
- [2] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Erik van der Goot, and Bruno Pouliquen. Opinion Mining on Newspaper Quotation. In *proceedings IEEE/WIC/ACM International Joint Conferences volume3*, Milan, Italy, 2009.
- [3] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment Analysis in the News. In *Proceedings of LREC*, Valletta, Malta, 2010.
- [4] 小林のぞみ, 乾健太郎, 松本裕治. 意見情報の抽出/構造化のタスク仕様に関する考察. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2006, No. 1, pp. 111–118, 1 2006.
- [5] 赤峯亨, 宮森恒, 加藤義清, 中川哲治, 乾健太郎, 黒橋禎夫, 木俣豊. 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, 2008.
- [6] 宮森恒, 赤峯亨, 加藤義清, 兼岩憲, 角薫, 乾健太郎, 黒橋禎夫. 情報の信頼性分析に向けた評価データおよびプロトタイプシステム WISDOM. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2007, No. 76, pp. 103–108, 7 2007.
- [7] 吉岡真治, 神門典子, 関洋平. 複数国の新聞サイトを比較分析する NSContrast の実験的分析. 情報処理学会研究報告 情報基礎とアクセス技術研究会報告, pp. 1–8. 情報処理学会, 2011.
- [8] 中野正寛, 洪木英潔, 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, pp. 496–499. 言語処理学会, 2011.
- [9] 河原大輔, 黒橋禎夫, 乾健太郎. 主要・対立表現の俯瞰的把握: ウェブの情報信頼性分析に向けて (情報分析・要約 (テーマセッション 1)). 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2008, No. 67, pp. 49–54, 7 2008.
- [10] 関洋平, Evans David Kirk, 神門典子, 青野雅樹. 意見保有者の判別: NTCIR 意見分析タスクにおける試み (「主観表現処理の最前線」シンポジウム). 電子情報通信学会技術研究報告. NLC,

言語理解とコミュニケーション, Vol. 107, No. 480, pp. 13–18, 1 2008.

- [11] 永井隆広, 石下円香, 中野正寛, 松本拓也, 洪木英潔, 森辰則. 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. 言語処理学会第 18 回年次大会, pp. 1051–1054, 2012.