Vol.2009-DBS-149 No.16
2009/11/21

論点に対する極性に注目した ニュース記事からの編集意図の抽出手法

青木伸也[†1] 湯本高行[†1] 角谷和俊[†2]
新居学[†1] 高橋豊[†1]

ニュースは同じトピックでも新聞社ごとに報道内容に違いがある．これは新聞社ごとに様々な論点に対しての見解が違うからで，1 つの新聞だけ読んだ読者は，他の新聞では異なる見解が示されているにも関らず，その 1 つの新聞と同じ見解を持ってしまう恐れがある．そこで，あるトピックについての新聞記事の集合から新聞社ごとの編集意図を抽出する手法を提案する．各見解は論点と極性（ポジティブ/ネガティブ）からなるとし，その集合として編集意図をモデル化する．抽出の際には，見解を述べていると考えられる見解文に注目し，精度よく編集意図を抽出することを目指す．ユーザは可視化された編集意図を見ることで各社の見解の違いを比較することができる．

Extracting Author Intention based on Polarity for Argument Point from News Articles

Shinya AOKI,[†1] Takayuki YUMOTO,[†1]
Kazutoshi SUMIYA,[†2] Manabu NII[†1]
and Yutaka TAKAHASHI[†1]

When the several authors report the same news topic, reported facts are often different by the author. It is because each author has his own observation about various points of the news topic. If users read newspapers of only one author, they obtain biased understanding about the news topic. In this paper, we propose the method for extracting author intentions. We model author intentions as sets of observations consisted of (argument-point, positive/negative, strength of observation). In our proposed method, we use sentences which often express observation to extract author intention in high accuracy. Users can compare some authors by looking at visualized author intention.

1. はじめに

世の中の出来事は新聞などで報道されるニュースを通して理解されることが多いが，これらのニュースの中には，見る者によって意見や印象が異なる出来事も存在する．そして，そのようなニュースは，同じトピックについての報道でも新聞社によって報道内容に違いがある．これは，新聞社ごとにトピックに関係するより小さな事柄に対しての見解が違うからである．例えば，2005 年 1 月にあった「特定外来生物問題」に関連する記事中にあった,「ブラックバスは日本の生態系にとって大きな脅威である」という文と,「バス釣りは釣り具業界などに 1 千億円の恩恵をもたらしている」という文では，ブラックバスに対する印象が大きく異なる．

一般に読者は 1 つの新聞の報道だけをみて，そのトピックを理解する．しかし，それではそのトピックに対して，選んだ新聞報道と同じ見解をもってしまう恐れがある．他の新聞が違った見解を示しているにも関わらずである．野中は，高校生，大学生を対象とした実験であるが，新聞のメッセージと読者の意見形成には密接な関係があるとしている[1)]．どの新聞記事が正しいのかはともかくとして，そのトピックについて深い理解を得るためには，各新聞社の報道を比較する必要がある．比較するためには，各社の新聞を読めばよいが，これは大変面倒な作業である．

そこで，あるトピックについて複数新聞社の意図を比較するために，各新聞社の意図を表す編集意図を抽出する手法と，それを可視化する手法を提案する．我々は編集意図を様々な論点に対する見解の集合としてモデル化する．各見解はある現実の出来事を編集することで示されるため，我々はこの集合を編集意図と呼んでいる．また，見解は論点と極性の組でモデル化する．論点は何に対して言及しているかであり，極性はそれに対してポジティブな見解か，ネガティブな見解かということを表す．すなわち，編集意図が表現するのは各新聞社が各論点に対してのポジティブな見解，ネガティブな見解をどれだけ持っているかということである．

各新聞社の編集意図はどの論点に対して言及しているかという点で異なっており，また同じ論点に対してでもポジティブな見解を多く持つ場合と，ネガティブな見解を多く持つ場合

†1 兵庫県立大学工学研究科
Graduate School of Engineering, University of Hyogo
†2 兵庫県立大学環境人間学部
School of Human Science and Environment, University of Hyogo

Vol.2009-DBS-149 No.16
2009/11/21

とで異なる．そのため，編集意図を適切に可視化することで，複数新聞社の報道がどう違うかを容易に比較できると考えている．

本稿では，まず，2 節で研究概要について述べ，3 節で関連研究について述べる．そして，4 節で手法について述べ，5 節で実験について述べる．最後に 6 節でまとめと今後の課題を述べる．

2. 研究概要

本節では研究概要について述べる．本研究で編集意図抽出の対象とする新聞記事，本提案手法における見解文の定義，また編集意図モデルについて，以下の節で詳しく述べる．

2.1 対象とする新聞記事

本研究の対象とする新聞記事は「社会」,「ビジネス」,「政治」,「国際」のいずれかについてのトピックに関するもので，ある程度大きな話題になったトピックに関するものである．

「社会」,「ビジネス」,「政治」,「国際」に対象を絞るのは，これらをテーマとするトピックは他のテーマに比べて，多様な見解が存在すると考えられるからである．また，ある程度大きな話題になったトピックしか扱わないのは，本提案手法では少量の記事から各新聞社の編集意図の違いを表現するのは難しいと考えるからである．

本研究では対象とする新聞記事と同じトピックでも「社説」や「コラム」は対象としない．これらには各新聞社の主張がはっきりと述べられているため，読者はそれがトピックに対する一意見だと認識していると考えられる．本提案手法が対象とする記事は，事実に織り交ぜて見解が述べられている記事である．このような記事は事実が述べられている分，読者も鵜呑みにしがちであるため，本提案手法により編集意図を抽出する必要があると考えている．

2.2 見解文

新聞記事には事実だけを述べた文と見解を述べた文が混在している．本提案手法では，見解を述べた文（見解文）のみに注目することで，精度よく編集意図を抽出する．本稿における見解文の定義を表 1 に示す．また，それぞれの場合に極性をどう判断すべきかを合わせて示す．

表 1 に示したもののうち，最後の 1 つ以外は，乾らが「意見」に含まれるとしているものである[2]．最後にあげた世の中の動きを記述したものには次のような文がある．

- 〜という意見（人，組織）が多い
- 〜という意見（人，組織）もある
- （国，組織）では，〜

表 1 見解文の種類

見解文の種類	極性の判断
評価を記述するもの	評価により P または N
要望，要求，提案の表明	P
不安，懸念，不満，満足等の感情を表すもの	感情により P または N
認識，印象を述べるもの	認識，印象により P または N
賛否の表明	賛ならば P，否ならば N
（新聞社の見解を裏付けるような）世の中の動きの記述	P

このような文は事実を述べているようにも見えるが，新聞社の編集意図を抽出するに当たっては重要な文だと考えられる．

2.3 編集意図モデル

我々は，あるトピックについての新聞社の編集意図はそのトピックについての様々な論点についての見解の集合だと考える．新聞記事中には各論点についての見解文が存在し，それらを総合することで編集意図を表すことができると考えている．見解文が表わすのはどの論点にどのように（ポジティブ／ネガティブ）言及していて，それらの見解がどの程度強く述べられているかである．これらを考慮し，編集意図を (1) 式のようにモデル化する．

$$intention = \{(pt_i, P, stren_i)|i = 1 \cdots n\} \cup \{(pt_j, N, stren_{n+j})|j = 1 \cdots n\} \quad (1)$$

pt_i は論点を，P，N はそれぞれポジティブ／ネガティブを表し，$stren_i$ は見解の強さを表す．同じ論点でも P/N の違いによって異なる見解として扱い，それぞれに見解の強さが存在するものとする．編集意図は論点，極性，見解の強さの 3 つ組の集合として表す．

各新聞社の報道で異なるのはこれらの点であるので，編集意図を用いることで新聞社の意図を表現することが可能だと考えられる．また，編集意図を比較することで各新聞社の報道の違いがわかると考えられる．

3. 関連研究

ニュース記事間の言及内容や見解の違いに注目したものとして以下の研究がある．

北山らは映像ニュースからテキストニュースを検索する比較ニュース検索方式を提案している[3]．北山らはメディア間でのニュースの構成の違いに注目している．映像ニュースとテキストニュースが相補的な関係にあるとして，異メディアの相補的なニュースを検索する方法を提案している．筆者は 1 つの新聞社だけでは見解に偏りがあると考え，複数の新聞社を比較することで深く理解できると考えている．その意味で新聞社の違うテキストニュース間

にも相補的な関係があるといえる．

灘本らは，発信国間で同じトピックに関するニュースに視点の違いがあると考え，B-CWB(Bilingual Comparative Web Browser) を提案している[4]．B-CWB では 2 カ国のニュース記事を比較し，差異情報を発見することができる．筆者は，トピックについてのより多くの記事から見解を抽出することで，新聞社間の違いが明確に表れると考えている．

吉岡らは，ニュース発信者の視点に注目し，トピックの全体像を構築する手法を提案している．筆者もニュース発信者の視点は重要と考えている．見解にのみ注目し，編集意図を抽出することで，よりニュース発信者の視点というものが把握しやすいものになると考えている[5]．

4. 編集意図の抽出手法

本稿では，複数新聞社のあるトピックに関する新聞記事集合が取得されているものとして，新聞記事集合からそれぞれの新聞社の編集意図を抽出する手法を提案する．記事中の見解文のみに注目し，その論点と極性から編集意図を導出する．編集意図抽出の処理手順は次のとおりである．

(1) **見解文抽出** 記事から見解文のみを抽出する
(2) **極性の判定** 見解文の極性を判定する
(3) **論点の抽出** 見解文の集合から論点を抽出し，見解文を各論点に分類する
(4) **編集意図の導出** 見解文の論点と極性に基づいて編集意図を構築する

上記のように手順を示したが，極性の判定と論点の抽出は処理が独立しているため，順序が逆でも構わない．それぞれの処理について以下の各節で述べる．

4.1 見解文の抽出

記事からの見解文の抽出処理は，記事中の文章を単位文に分割する処理と単位文ごとに見解文であるか否かを判定する処理の 2 段階で行う．見解文を判定する処理では SVM を用いて機械学習を行い，分類する．4.1.1，4.1.2 節でそれぞれ詳しく述べる．

4.1.1 記事の単位文への分割

見解文であるかどうかを判定するのは基本的には記事中の文単位である．しかし，人の発言の引用（かぎ括弧 “「」” で囲まれた文）があったり，複文（読点 “、” で分割できる）があったりするため，以下のような分割処理を行う．

- **記事の分割** 記事を句点 “。” で分割する．ただし，引用文内（“「」” 内）の句点は無視する．
- **複文の分割** 文中の読点 “、” で分割する．ただし，分割後の助詞・助動詞を除いた文末に「動詞」,「形容詞」,「形容動詞」が現れる場合のみ，つまり述語が存在する場合のみ分割する．
- **引用文の分割** かぎ括弧 “「」” で囲まれた文を抜き出して 1 文とする．また，それに対して，文単位の分割や複文の分割を行う．

図 1 に処理の概要を示す．図に示す例では，最終的に b 以外の文が見解文かどうかの判定対象となる．引用文は他の文と重複することになるが，新聞社が他人の言葉を借りて見解を述べていることも考慮して見解文の判定の際の単位文としても扱うことにする．

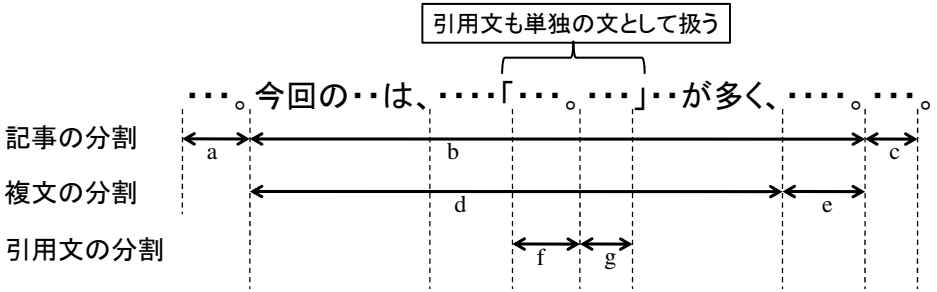

図 1 分割処理

4.1.2 見解文の判定

4.1.1 節の処理で分割したものを，SVM を用いて機械学習を行い，単位文ごとに見解文であるか否か判定する．見解文であるか否かは文末に注目して抽出する．我々は，見解文は文末に特徴が現れると考えた．松本らは文末表現に注目してウェブページの主観・客観度を判定しているが，見解文も主観的な文に近いといえる[6]．そのため，松本らの手法を参考に文末を「**文の末尾から遡り，自立語が出現するまでを，自立語を含めて文末とする**」と定義する．松本らも述べているが，文末に注目することでトピックへの依存性は低くなると考えられる．

SVM に与える素性としては，文末中の各単語の出現回数を用いる．ただし，自立語である単語は種類が多く単語そのものがそれほど分類に重要でないと考えられるため，単語そのものではなく品詞情報（名詞，動詞など）を素性とする．例外として,「思う」,「考える」,

「感じる」,「願う」の 4 動詞は共通性も高く，見解文判定に重要だと考えられるので，単語そのものを素性とする．また，文末以外でも副詞の存在も見解文判定に重要だと考えられるので，文全体での副詞の出現回数も素性として用いる．

4.2 極性の判定

4.1 節の手法で抽出した各見解文に対して，極性が P（positive）であるか，N（negative）であるか，または E（neutral）であるかを判定する．判定には辞書を用いる．極性が E と判定された見解文は編集意図の導出の際には用いない．

4.2.1 節で見解文の極性判定について，4.2.2 節で用いる辞書について述べる．

4.2.1 見解文の極性判定

辞書参照により見解文に含まれる単語の極性を取得，または極性反転子であるかを判定し，それらを合わせて見解文の極性を決定する．対象とする単語の品詞は「名詞」,「形容詞」,「形容動詞」,「動詞」である．以下で具体的に述べる．

見解文の極性は，単語から判断できる極性を極性反転子の出現回数分だけ反転したものとなる．極性反転子は単語単体では極性を持たないが，他の極性を持った単語と組み合わさることで全体として逆の極性をもつものである．例えば「ごみ」という名詞は N の極性を持つが,「ごみが少ない」という文は極性反転子である「少ない」という形容詞のため，文全体として P の極性を持つ．乾らは plus，minus 属性をもつ単語として pn 演算子を定義しているが，このうち minus 属性を持つものが極性反転子である[7].「ごみが少なくない」という文には，極性反転子が 2 つ現れる（「少ない」と「ない」）ので，文の極性は N である．

また，単語から取得できる極性として，矛盾する極性が得られる場合が考えられるが，その場合は文の末尾に近いほうの単語の極性を基準とする．これは見解文の極性の決定には述語が重要で，述語が極性を持つ場合は主語の極性は無視できると考えられるからである．

見解文中のどの単語も P または N の極性を持たない場合，見解文の極性を E と判断し，編集意図の導出には利用しない．

4.2.2 単語極性辞書・極性反転子辞書

極性の判定には以下の辞書を使用する．

- 単語極性辞書
- 極性反転子辞書

単語には 1 単語だけで P（N）の極性をもつものがあるが，そういった単語を集め，極性とともに記録した辞書が単語極性辞書である．単語極性辞書として以下の 2 つの既存辞書を使用する．

- **名詞，形容動詞用の辞書** 日本語評価極性辞書（名詞編）[8]
- **形容詞，動詞用の辞書** 単語感情極性対応表[9][10]

日本語評価極性辞書は人手によるチェック済みの辞書で名詞が収録されている．また，形容動詞も形容動詞語幹名詞として収録されている．単語感情極性対応表は日本語評価極性辞書に収録されていない，形容詞，動詞の極性判断の辞書として用いる．極性反転子辞書は著者らが作成したものを使用する．作成に当たっては，一般的な対義語・反対語リストを参考に作成した[11][12][13].

4.3 論点の抽出

4.1 節の手法で抽出した見解文集合から論点を抽出する．最終的にユーザに提示するには重要な論点だけを示すほうが良いと考えられる．また，各論点は単語集合で表すのが妥当と考える．そのため，まず各見解文から論点を構成する論点要素語を抽出し，できるだけ少ない論点で多くの見解文をカバーするような論点集合を見つけることで，重要な論点を抽出する．

それぞれの見解文には，論点要素語抽出の際に抽出した論点要素語集合に論点中のすべての論点要素語を含むかどうかで，発見した論点集合のいずれか，または複数の論点を対応付ける．

4.3.1 論点要素語の抽出

各見解文から名詞や名詞の連続など名詞句のパターンに一致する位置に出現する単語を論点要素語として抽出する．論点要素語の抽出には MeCab を用いる[14].

名詞句のパターンを表 2 に示す．表 2 中では，MeCab で「名詞，形容動詞語幹」として扱われる単語を形容動詞,「名詞，サ変接続」として扱われる単語をサ変接続名詞としている．表 2 は佐々木らの名詞句の定義を参考にしたものである[15]．名詞句のパターンに一致する位置に出現する単語が，それぞれ別々に論点要素語として抽出される．例外など詳細な設定を以下に述べる．

- 数詞は抽出しない
- 平仮名のみの単語は抽出しない
- 接頭詞は名詞として扱う
- 1 文字の単語は抽出しない

見解文からの論点要素語の抽出以外に，局所文脈という概念を導入し，新聞記事中での前 2 文中の単語も素の見解文に出現する単語として抽出する．ここでいう文は本提案手法で扱う単位文である．局所文脈は注目する文に意見の対象となる語が存在せず，周辺文に対象語

表 2 名詞句のパターン

パターン	例
(a) 名詞	情報
(a) 名詞の連続	情報／公開，独占／禁止／法
(b) 名詞句＋の＋名詞句	医療／の／質
(c) 名詞句＋（と｜や｜・）＋名詞句	水／と／油
(d) 形容動詞語幹＋な＋名詞句	重要／な／情報
(e) 形容詞＋名詞句	軽い／怪我
(f) 動詞＋名詞句	調べる／方法
(g) 名詞句＋（が｜を｜は｜の｜も）＋サ変動詞語幹	情報／の／公開

が存在する場合を考慮した概念で，岡本らが提案している[16]．

4.3.2 論点集合の発見

4.3.1 節で抽出した論点要素語から重要な論点の組み合わせを発見する手法について述べる．

まず，抽出された論点要素語と抽出元の見解文の関係から，各見解文に対応する論点要素語集合を考える．各見解文の論点要素語集合と各論点要素語をそれぞれ，相関ルールマイニングにおけるトランザクションとアイテムと考え，頻出飽和アイテム集合を抽出する．頻出飽和アイテム集合の抽出には LCM を用いる[17]．アイテム集合を頻出かどうか判定するための最小支持度 min_{supt} は (2) 式で計算する．

$$min_{sup} = \frac{|S|}{2 \cdot P_{num}} \tag{2}$$

式中の S は見解文集合を表し，P_{num} は抽出される最大の論点数であり，パラメータとして設定しておく，頻出飽和アイテム集合の抽出は論点を表すには妥当ではない論点要素語を削除するためと，複数単語で論点を表すようなものを発見するために行う．

次に，発見された各頻出飽和アイテムを論点候補と考え，全体の論点候補集合を CP とする．CP の部分集合 CP_{sub} を組み合わせ候補とし，できるだけ少ない論点候補の組み合わせで多くの見解文をカバーするような，CP_{sub} を探す．すなわち，(3) 式の評価関数の値ができるだけ大きくなる組み合わせ候補 CP_{sub} を探索する．

$$eval(CP_{sub}) = \frac{(\text{網羅度})^2}{(\text{論点数}) \cdot (\text{標準偏差})} = \frac{(|\bigcup_{cp \in CP_{sub}} \{s|cp \subset T_s\}|/|S|)^2}{(|CP_{sub}|+1) \cdot (sd(CP_{sub})+1)} \tag{3}$$

式中の cp は論点候補，T_s は見解文 s の論点要素語集合を表す．論点数や標準偏差に 1 を足しているは，これらの値が 0 になり全体の値が無限大となるのを防ぐためである．$sd(CP_{sub})$ は (4) 式で計算する．

$$sd(CP_{sub}) = \sqrt{\frac{\sum_{cp \in CP_{sub}} |\{T_s|cp \subset T_s\}| - mean(CP_{sub})}{|CP_{sub}|}} \tag{4}$$

$$mean(CP_{sub}) = \frac{\sum_{cp \in CP_{sub}} |\{T_s|cp \subset T_s\}|}{|CP_{sub}|} \tag{5}$$

分散を評価関数に加えたのはあまりに大きさの違う論点の組み合わせが抽出されないようにするためである．

評価関数が大きくなるような論点候補の組み合わせの探索には山登り法を用いる．ランダムに複数の解を生成し，それぞれ近傍解を調べながら探索を行うことで評価関数が大きくなるような論点候補の組み合わせを見つける．解を論点候補数と同じ数のビット列で表し，各論点候補に対応する位置のビットの値（0/1）で解中に各論点候補を含むか否かを表したとき，ハミング距離が 1 である解を近傍解とする．

ある程度探索を繰り返し，探索した中で評価関数が最大となる解，すなわち論点候補集合を論点集合として抽出する．

4.4 編集意図の導出

4.1，4.2，4.3 節で抽出した見解文，極性，論点を用いてそれぞれの新聞社ごとに編集意図を導出する．論点集合を PT として，(6) 式で編集意図を導出する．

$$intention = \{(pt, pl, |\{s|Point(s) \ni pt, Polarity(s) = pl\}|)|pt \in PT, pl \in \{P, N\}\} \tag{6}$$

$Point(s)$ は見解文 s に対応付けられた論点の集合で，$Polarity(s)$ は見解文 s の極性を表す．編集意図中の各見解の強さには各論点，極性の見解文数を用いる．ただし，1 つの見解文が複数の論点についての見解文であることを許容し，計算しているため，編集意図中の見解の強さの総和と全見解文数は一致しない．

5. 実　　験

本提案手法における見解文抽出の精度を確かめるための実験と，論点抽出で必要な論点を抽出できているかを確かめる実験を行った．また，論点が同じ場合でも極性が異なる見解文の例を示す．以下，それぞれの節で述べる．

5.1 見解文抽出の実験

本提案手法における見解文抽出の精度を調べるため，10-fold cross validation による評価を行った．実験に用いた正解データセットは 4 人の作業者に協力してもらい作成した．

以下，5.1.1 節で正解データセットについて，5.1.2 節で 10-fold cross validation による評価と考察について述べる．

5.1.1 正解データセット

表 5.1.1 にあげるトピックについての新聞記事について，各文が見解文であるか否かのラベルを付与することで見解文の正解データセットを作成した．作業者は 4 人で 2 人ずつのグループに分かれ，それぞれ表 5.1.1 に示す A グループ，B グループのトピックについて，記事を 4.1.1 節の方法であらかじめ文単位に分割したものにラベル付けを行ってもらった．作業者には 2.2 節の定義を伝えた．また，作業者にははっきりと見解文だと分かるものについてラベル付けしてもらうよう依頼した．

表 3 正解データセットの作成に用いた記事

	トピック	期間	全記事数	朝日記事数	毎日記事数	読売記事数	全文数
A	JR 福知山線 運転再開	2005-6/13-6/27	39	6	14	19	2084
A	アスベスト 対策	2005-7/15-8/6	124	34	48	42	3317
A	靖国問題	2005-6/14-6/28	55	17	19	19	1296
B	人民元 切り上げ	2005-7/20-8/1	86	22	30	34	2281
B	北朝鮮 核実験	2005-5/5-5/19	49	12	21	16	928
B	個人情報保護法	2005-3/25-4/7	20	3	11	6	934

ラベル付け作業完了後，同じ記事について 2 人の作業者がラベルを付けているので，どちらかの作業者が見解文のラベルを付与しているものを見解文であるとして，結果の統合を行った．どちらか一方が見解文と判断していればよいとしたのは，作業の際にはっきりと見解文だと分かるものについて見解文のラベルを付けてもらうように依頼したためである．両方の作業者が分からないとしたものは正解データセットには加えていない．

作成した正解データセット中の見解文とそれ以外の文の割合を表 5.1.1 に示す．

5.1.2 10-fold cross validation による評価

作成した正解データセットを用いて，文末に注目し見解文を抽出することが妥当であるかを確かめるため，SVM を用いて 10-fold cross validation を行った．SVM は TinySVM を用い，RBF カーネルを選択した以外はデフォルトの設定で学習を行った[18]．

4.1.2 節の通りに素性を選択しデータを作成した結果，285 次元のベクトルデータとなった．10-fold cross validation を 10 回行った結果の平均値は精度が 74.9%，適合率が 69.2%，再現率が 54.8%であった．精度は SVM とテストデータの分類の一致率，適合率は SVM が見解文に分類したもののうちテストデータでも見解文と分類されているものの割合，再現率はテストデータにおいて見解文に分類されているもののうち，SVM が見解文に分類したものの割合である結果は十分とは言えないが，文末に注目して見解文を抽出することは妥当であると考えられる．

表 4 正解データセット中の見解文の割合

トピック	見解文数	それ以外の文数	見解文の割合
JR 福知山線 運転再開	653	1431	0.31
アスベスト 対策	897	2421	0.27
靖国問題	402	894	0.31
人民元 切り上げ	1092	1154	0.48
北朝鮮 核実験	477	416	0.51
個人情報保護法	365	559	0.39

まだ見解文の定義に曖昧な部分が残っているため，同じ作業者でもラベル付けに一貫性がない部分があるのではないかと考えられる．実際，作業者からラベル付けに一貫性がない部分があると思うという声が聞かれた．今後，見解文の定義から曖昧性を極力排除したり，作業者の数を増やし多数決で結果を統合したりするなど正解データセットを修正していく必要がある．

また，本稿では SVM の設定として RBF カーネルを選択した以外はデフォルトの設定で実験を行ったが，最適なカーネルパラメータを設定することで精度が改善する可能性がある．

5.2 論点抽出の実験

5.1.1 節で述べた正解データセットの見解文を用いて論点の抽出を行った．正解データセットの作成に協力してもらった作業者 4 人に，記事を読んでどんな論点が印象に残ったかアンケートを取り，抽出結果と比較した．これは重要な論点は印象に残りやすいだろうという考えから行ったものである．表 5 から表 10 にそれぞれのトピックの抽出した論点を示す．(2) 式で使用する最大論点数として，20 を指定した．

各表中で*が前置されている論点は作業者の印象に残っていた論点と内容的に同じものである．例えば，表 5 にある論点「通勤 JR」や「通学」には，通勤や通学のためには乗らないと仕方がないといったような見解文があった．また,「余裕 ダイヤ」にはダイヤに余裕がなかったこと,「装置」には新型自動列車停止装置を設置していなかったことに対する見解文があった．これらは，作業者の印象に残っていた論点であるので，重要な論点が抽出できていると考えられる．抽出できなかった論点としては「脱線したものと同じ型の車両は二度と

Vol.2009-DBS-149 No.16
2009/11/21

使わない」や「ハード面は良いが運転手の意識が大切」といった割と小さな論点があった．しかし，このような論点も印象に残ることがあるということは考える必要がある．また,「向上」,「仕方」など論点としてふさわしくないものもある．これについても考える必要がある．

表 8 では，多くの論点に「切り上げ」が存在するが，これはトピック全体を表現する単語なので，論点には必要がないものだと考えられる．実際に，そのような論点を見ても「切り上げ」が論点の内容を推測する助けにはならない．表 9 では「北朝鮮」，表 10 では「情報」がそれにあたる．

表 5 「JR 福知山線 運転再開」の結果

電車 運転
説明 遺族 JR
速度 遺族
宝塚 JR
会社 再開
*通勤 JR
思い 事故
*余裕 ダイヤ
*通過 現場
*複雑 再開
*最高 速度
息子
時分
*改正 ダイヤ
向上
*通学
不通
言葉
回復
仕方
誠意 誠心
*装置
社長

表 6 「アスベスト 対策」の結果

周辺 被害 対策
*労災 アスベスト
労働 石綿
従業 住民
*禁止 石綿
*代替 石綿
多く
連絡
個人
*申請
現場
センター
*自治体
情報
記者
*施設
治療
*法案

表 7 「靖国問題」の結果

小泉 施設
小泉 靖国
記者 靖国
*考え 追悼 施設
*考え 韓国
発言 遺族
判断 靖国
*大統領 韓国
*自身 日本
過去 歴史
長官 検討
政治 参拝
幹部 参拝
*中止 参拝 首相
誤解 施設
国内
民主党
確認
*批判 首相
宗教 施設
世論
今後
説明
外国
意味
気持ち

表 8 「人民元 切り上げ」の結果

日本 為替
日本 通貨
*変動 影響
改革 為替 中国
見方 相場
*世界 日本 中国
米国 経済
*企業 市場
銀行 通貨
当局 制度
動き 人民元
バスケット ドル
取引 ドル
*歓迎 切り上げ
外国 為替 切り上げ
国際 切り上げ
懸念 中国
赤字 切り上げ
時期 切り上げ
マイナス 中国
会長 切り上げ
結果
一時
混乱
資金

表 9 「北朝鮮 核実験」の結果

中国 米国
*発表 使用 北朝鮮 燃料 済み
外交 協議
関係 協議
*制裁 国連 実験
政権 ブッシュ 協議
宣言 核兵器 北朝鮮
社会 国際 実験 北朝鮮
日本 国連
6 月 米国
技術 北朝鮮
実施 実験 北朝鮮
*兆候 準備 実験
強い 協議
偵察 政府 衛星
*今回 燃料
*解決 問題 協議
具体 北朝鮮
停止 4 月
会談 北朝鮮
*反応 実験
計画 北朝鮮
トンネル 北朝鮮
*完了 燃料 取り出し
条約
合意
ワシントン
国務 北朝鮮
必要 核兵器
立場
事務 局長
専門

表 10 「個人情報保護法」の結果

*患者 病院 情報
利用 目的 情報
*国家 発表
法律 個人 情報
対策 企業
*国民 合格 医師
対象 情報
電話
*保険
業界
請求
*委託
番号
営業
第三者
責任 情報
以降

5.3 同じ論点で極性の違う見解文の例

5.2 節で論点抽出した結果から，論点が同じでも極性が異なる見解文の例を表 11,12 に示す．表から同じ論点でも極性の異なる見解文を別々に扱うことが重要であると分かる．

表 11 トピック：「JR 福知山線 運転再開」，論点：「通勤 JR」

見解文	極性
これから通勤が楽になる	P
短時間にのりかえなしで大阪へ行ける JR はやっぱり便利	P
ありがたいが，これまでの過密ダイヤ … 運転士がゆとりを持てるかどうか	N
JR に乗らざるを得ない時も来ると思うので、複雑だ	N

表 12 トピック：「アスベスト 対策」，論点：「代替 石綿」

見解文	極性
代替品の開発普及の 重要性を強調	P
… 代替品に変えることで … 石綿繊維の大気中への放出を完全に抑制できる	P
ガラス繊維やロックウール（岩綿）を検討したが … 安全性にも不安が残った	N
安易な代替品は望ましくない	N

6. まとめと今後の課題

本稿では複数の新聞社の報道を比較するために編集意図を抽出する手法を提案した．編集意図を見解の集合として，見解を論点と極性の組としてモデル化することで，新聞社間の報道の比較を編集意図間の比較をすることで行うことができると考えている．

本提案手法では編集意図抽出に見解文のみを用いることで，意図を表す際にノイズとなる文を排除する．また，論点とそれに対する極性に注目して編集意図をモデル化し，導出する．実験では，SVM を用いた文末に注目した見解文抽出の精度を確認する実験と，論点抽出結果の人の印象に残った論点との比較をした．見解文抽出の実験では，十分な精度とは言えないが，文末に注目して見解文を抽出することはある程度妥当だという結果を得ることができた．今後は正解データセットの修正や SVM の最適なカーネルパラメータの探索を行う必要がある．論点抽出の実験では，重要な論点が抽出できていることが確認できた．今後は必要のない論点を排除し，精度を向上させることが必要である．また，本稿では論点が同じ場合でも極性が違う見解文を示すことで，極性を考慮することの重要性を示すにとどまったが，極性判定についても実験を行うことが必要である．

その他今後の課題として，本稿ではあるものとしていた新聞記事集合を取得する方法の考案，論点抽出の際の各パラメータや式の細かな分析が挙げられる．

謝辞 本研究の一部は，NICT 委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」によるものです．ここに記して謝意を表すものとします．

参 考 文 献

1) 野中博史：報道による意見形成効果 : NIE への指針, 宮崎公立大学人文学部紀要, Vol.13, No.1, pp. 261–274 (20060320).
2) 乾孝司, 奥村学：テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp. 201–241 (20060710).
3) 北山大輔, 角谷和俊：コンテンツ構成要素の順序特性に基づく比較ニュース検索方式, 電子情報通信学会研究報告, Vol. 107, No. 131, pp. 277–282 (2007).
4) 灘本明代, 田中克己：B-CWB: 類似コンテンツの視点差異情報を同時提示する多言語 Web ブラウザ, 日本データベース学会 Letters, Vol.2, No.2, pp. 13–16 (2003).
5) 吉岡由智, 湯本高行, 田中克己：ニュースの視点の抽出によるマルチメディアニュースアーカイブの利用, 電子情報通信学会研究報告, Vol. 2005, No.67, pp. 415–420 (2005).
6) 松本章代, 小西達裕, 高木朗, 小山照夫, 三宅芳雄, 伊東幸宏：文末表現を利用したウェブページの主観・客観度の判定, 第 1 回データ工学と情報マネジメントに関するフォーラム (2009).
7) 乾孝司, 乾健太郎, 松本裕治：出来事の望ましさ判定を目的とした語彙知識獲得, 言語処理学会第 10 回年次大会, pp. 91–94 (2004).
8) 日本語評価極性辞書, http://cl.naist.jp/~inui/research/EM/sentiment-lexicon.html.
9) 単語感情極性対応表, http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html.
10) 高村大也, 乾孝司, 奥村学：スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌ジャーナル, Vol.47, No.2, pp. 627–637 (2006).
11) 形容詞（けいようし）の反対語（はんたいご）, http://china.web.infoseek.co.jp/japanese/b-kei-hantai.pdf.
12) 対義語・反対語辞典, http://hanntaigo.main.jp/.
13) 浜島書店-対義語 (表), http://www.hamajima.co.jp/kokugo/gyakubiki/contents/taigigo.shtml.
14) MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/.
15) 佐々木千晴, 藤井敦, 石川徹也：意思決定支援のための主観情報マイニング, 言語処理学会第 12 回年次大会発表論文集, pp. 77–80 (2006).
16) 岡本和剛, 本田徹也, 江口浩二：意見文検索のための言語モデルにおける局所文脈スムージング, 情報処理学会研究報告, Vol. 2009-FI-95, No.16, pp. 1–7 (2009).
17) 宇野毅明と有村博紀による公開プログラム（コード）, http://research.nii.ac.jp/~uno/codes-j.htm.
18) TinySVM: Support Vector Machines, http://chasen.org/~taku/software/TinySVM/.