

Web ニュースの閲覧記事に対する対立記事抽出手法

大原 正章^{†,a} 真下 遼^{†,b} 灘本 明代^{†,c}

† 甲南大学大学院自然科学研究科知能情報学専攻

a) *m1524001@s.konan-u.ac.jp* b) *hr-x7type@yahoo.co.jp* c) *nadamoto@konan-u.ac.jp*

概要 Web ニュース記事を読む際、1つの記事を読んだだけでは内容の重要性を把握できない場合がある。そこで我々は、ユーザの閲覧している記事と対立している記事を提示することにより、重要性を把握する手助けとなると考え、この対立している記事を自動で検索し提示する手法を提案する。本研究では、ニュースの主題は異なるが内容の似ている記事を対立記事とする。具体的には、ユーザの閲覧している記事から主題となる単語(主題語)と記事の観点(記事アスペクト)を抽出し、この主題語と対立する単語が主題となっている記事に対して閲覧記事と同じ記事アスペクトを持つ記事を対立記事として抽出して表示する手法を提案する。

キーワード ニュース記事, Web, 対立記事

1 はじめに

近年、インターネットの普及により Web 上のニュースサイトを利用するユーザが増加している。Web ニュースはいつでもリアルタイムにニュースを取得することが出来るため、Web ニュースを利用することは最新ニュースを取得する有用な手段の一つであると考えられる。しかしながら、Web ニュースを読む際、1つの記事を読んだだけでは内容の重要性を把握できない場合がある。例えば、ある企業の赤字に関する記事を閲覧した時に、その企業の事は知っていてもその業界についての知識が無い場合、閲覧記事がどのくらい重大であるのかを理解することは容易ではない。このような場合、その企業のライバル企業の収支に関する情報と比較することにより、閲覧中の記事の重要性を把握することが可能であると考えられる。

一方、現在の Web ニュースには、その記事に対する関連記事のリンクが多数存在する。しかしながら、その関連記事の多くは閲覧している記事の過去に報道された記事である場合や、閲覧記事内に出現しているキーワードに関する記事である場合がほとんどである。このような関連記事は、経過情報を記載したニュースやニュースの主題が同じというだけで内容に関連しないニュースであることが多い。そのため、関連ニュースを閲覧した場合においても、元のニュースの重要性を理解することは困難である場合が多数ある。

そこで我々は、閲覧記事と対立(ライバル)関係にある記事を提示することにより、その記事の重要性を理解することが可能であると考え、閲覧記事の対立関係にある記事を抽出し提示する手法を提案する。本論文では、対立関係にある記事を対立記事と呼ぶ。この時、ニュース記事には複数の観点がある場合が多数ある。例えば、

「A社の最終決算が10億円の赤字」に関する記事では、最終決算が10億円の赤字であることがどのくらいすごいかわからない時、そのニュースの観点として、「最終決算」と「赤字」が考えられる。「最終決算」を観点とした場合、「B社の最終決算は2億円の黒字」を示した記事が対立記事となり、「赤字」を観点とした場合、B社の記事は黒字のため対立記事とはならず「C社は中間決算で2億円の赤字を出した」が対立記事となる。このように、最終決算では他社はどのようになっているかがわかる観点と、赤字がどのくらいすごいかがわかる観点がこの記事には含まれていると考えられる。そこで我々は記事の観点に着目し、本研究ではユーザの閲覧記事の観点毎の対立記事を抽出し提示する手法を提案する。具体的には、ユーザが閲覧している記事から主題となる単語を抽出する。次にその記事の観点を抽出する。本論文では、記事の主題となる単語を主題語と呼び、記事の観点を記事アスペクトと呼ぶ。さらに、主題語と対立する単語である対立語を抽出し、対立語と記事アスペクトを含む記事に対して閲覧記事との類似度の最も高い記事を対立記事として抽出し、提示する手法を提案する。

2 関連研究

Web ニュース記事の理解を支援することを目的とした研究として、北山ら [1] は、映像ニュースとテキストニュースの比較のための質問生成の提案を行っている。本研究ではテキストニュースである Web ニュース記事のみを扱っている点で異なるが、一般に Web ニュース記事では理解に重要なことから先に書かれている点に着目して単語の重要度を決めており、本研究でもニュース記事の本文一段落目と二段落目以降で単語の重要度を付与している。池田ら [2] は、ニュース記事と blog 記事を対応付けることでニュースの理解を支援する手法を提案している。本研究では対立するニュース記事同士を対応

付ける点異なる。

ニュース記事の特徴語を抽出する手法として、田中ら [3] は、記事内で現れる人物や組織、場所、建造物などのエンティティをニュース記事の話題となる特徴語として抽出する手法を提案している。また、佐藤ら [4] は、ニュース記事において「政治」「スポーツ」「経済」などの分類を行い、更にクラスタリングすることで特徴語である話題語の抽出を行っている。本研究での特徴語として、ニュース記事の主題語とその記事の観点を示す記事アスペクトを抽出している点異なる。

文書比較を行う研究として、宮部ら [10] は、複数の新聞記事に対して機械学習することで、異なる記事において同じ内容を述べているかを特定する手法を提案している。本研究とは、機械学習を行わない点及び記事アスペクトに基づいた対立記事を抽出しているため異なる。

本研究では対立記事の抽出を行う際に対立語の抽出を行っているが、対立語を抽出する研究では、大島ら [5] らは、互いに対照的な関係性にある語の発見手法について並列助詞「や」で接続されている語をキーワードに対する関連語として抽出する手法を提案している。また、真下ら [6] は、漫才台本の自動生成において1つの対立語を用いて行うボケを生成する手法について提案している。佃ら [7] らは、認知度の近い語の関係から同意語を発見する手法を提案しており、Yahoo!ウェブ検索¹での検索結果数を認知度としている。本研究では上位概念と認知度を考慮したうえで複数の対立語を抽出する点異なる。

3 対立記事の提示手法

本研究では、ユーザの閲覧している記事の記事アスペクトに着目し、記事アスペクト毎に対立記事を抽出し提示する手法を提案する。以下と図1に提案手法の概要を示す。

- (1) ユーザの閲覧記事から主題語を抽出する。
- (2) ユーザの閲覧記事から記事アスペクトを抽出する。
- (3) 主題語に対してライバルとなる対立語を抽出する。
- (4) (2)で抽出した記事アスペクトと(3)で抽出した対立語との共起度を求め、ある記事アスペクトに対して最も共起度の高い対立語をその記事アスペクトとのペアとする。
- (5) (4)の結果の記事アスペクトと対立語のペアをクエリとした全ての記事を取得する。取得した全ての記事を対立記事の候補とする。
- (6) それぞれの対立記事の候補に対して記事アスペクトを取得し、閲覧記事の記事アスペクトとの類似

度を求め、最も高い類似度の記事を対立記事として決定し、提示する。

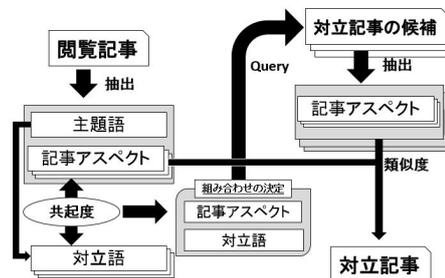


図1 提案手法の概要

本研究では、閲覧記事および対立記事を抽出するニュースサイトの対象は産経ニュース²とする。

3.1 主題語の抽出

一般にニュース記事は、タイトルでは記事として最も重要な情報を伝え、本文の1段落目ではタイトルの内容をさらに詳しく伝え、そして本文の2段落目以降では記事の内容に付随する事柄や背景などを詳細に伝えている。そこで本研究ではタイトル、本文の1段落目、本文の2段落目以降のように、段落の出現順で重要であると考え、ニュース記事内に出現する単語に対して出現位置を考慮した重みを付与する。また、ニュースの主題語は、そのニュースに頻出している単語で且つ固有名詞である場合が多いと考え、固有名詞を対象とする。

そこで、ある固有名詞*i*に対する重み S_i を式(1)に示す。 S_i がもっとも高い固有名詞を主題語とする。

$$S_i = \alpha \times tf_{il} + \beta \times tf_{im} + \gamma \times tf_{in} \quad (1)$$

ここで、タイトルの重みを α とし、タイトル位置*l*におけるある固有名詞*i*の単語の出現頻度を tf_{il} とする。同様に本文1段落目の重みを β 、本文の1段落目位置*m*におけるある固有名詞*i*の出現頻度を tf_{im} 、本文の2段落目以降の重みを γ 、本文の2段落目以降位置*n*におけるある固有名詞*i*の出現頻度を tf_{in} とする。重みについては、タイトル位置に一度しか出現しない単語より本文の1段落目位置に二度出現している単語のほうが重要であるが、本文の1段落目位置に一度しか出現していない単語でも本文の2段落目以降に二度出現している単語より重要であると考え、重み α 、 β 、 γ の値はそれぞれ0.5、0.3、0.1とする。

3.2 記事アスペクトの抽出

ニュースには、一つの記事に対して複数の記事アスペクトを持つ場合がある。例えば、「とある企業の最終決

¹Yahoo!検索. <http://search.yahoo.co.jp/>

²産経ニュース. <http://www.sankei.com/>

算が10億円の赤字」に関する記事では、「最終決算」や「赤字」が観点となる。また、記事内で赤字の原因が強調されている場合は、その原因も記事の観点になる場合があると考えられる。そこで、本研究ではユーザの閲覧記事からすべての記事アスペクトの抽出を行う。

本研究で提案する記事アスペクトは記事の特徴となる単語であることが好ましい。そのため記事アスペクトとなる単語も記事内では主題語と同じ程度に重要であると考え、式(1)を用いて固有名詞以外の名詞の重みを算出して閾値以上の全ての名詞を記事アスペクトとする。ここで、閾値は0.6とする。

3.3 対立語の抽出

我々の提案する対立語は、「東京」と「大阪」、「野球」と「サッカー」といったように、2つの語同士が互いに対照的な関係性にある語のことを指す。2つの語の関係性に着目すると「東京」と「大阪」は共に日本の都市、「野球」と「サッカー」は共に球技のように、対立関係にある語同士は共通の上位概念を持っていることがわかる。ここで、本研究では共通の上位概念を持つ語を対立語と定義する。また、「野球」の上位概念である球技に関して、同じ球技を上位概念に持つ兄弟語は「サッカー」だけではなく「フットサル」も挙げられるが、「サッカー」と「フットサル」では競技人口に大きな差があり、「野球」と同程度に認知されている「サッカー」の方が対立語として適切であると考えられる。そこで本研究では対立語を、主題語と共通の上位概念を持ち、且つ同程度の認知度を持つ語と定義し、対立語を抽出する。

3.3.1 対立関係の抽出

様々な概念辞書があるが新語にも対応するために、本研究では概念構造に Wikipedia のカテゴリ構造を用いる。例えば、東京では「日本の都市」や「就航地」など11語の上位概念が取得できる。このように上位概念は1つとは限らず複数個持つ場合がある。そこで、これらの上位概念のうち1つでも共通の上位概念を持つ語は主題語との兄弟語として取得する。

得られた兄弟語に関して、共通の上位概念を多く持つ兄弟語の方がより主題語と関連が強いと考えられる。また、どの上位概念が主題語と共通しているのかも考慮する必要がある。例えば、東京の上位概念である「作品」は120257語の兄弟概念を持ち、「日本の都市」は24語の兄弟概念を持つが、同じ上位概念でも少数の兄弟概念を持つ語の方がより重要性が高いと考えられる。そこで、主題語の上位概念にその下位概念の数を考慮した重みを付与し、取得した兄弟語と主題語の共通する上位概念の数だけ対応した重みを全て加算することで兄弟語の重みを求める。式(2)に主題語 s のある上位概念 U_s の重み $Sta(U_s)$ を示す。ここでは、上位概念 U_s が持つ下位概

表1 上位概念毎の兄弟概念数とその重み

上位概念 U_s	兄弟概念数 n	重み $Sta(U_s)$
日本の都市	24	11.7
就航都市	240	9.4
就航地	258	9.3
グループ	333	9.1
過去に実在した店舗	347	9
オリジナルアルバム	1100	7.8
曲	3717	6.7
シングル	19519	5
収録曲	35981	4.4
登場人物	95662	3.4
作品	120257	3.2

念数 n と主題語 s のすべての兄弟概念数 N_s を用いて以下の式で求める。

$$Sta(U_s) = \frac{\log n}{N_s} \quad (2)$$

式(2)を用いて東京の上位概念の重みを計算した値を表1に示す。兄弟概念の数が少ない上位概念ほど高い重みが与えられるのがわかる。

この上位概念の重みの値が上位10件の上位概念に対して全ての下位概念を取得し、主題語の兄弟語とする。次にそれぞれの兄弟語 b について主題語と共通する上位概念に対して式(2)を用いて兄弟語 b の重み $Rel(b)$ とし、以下の式を用いて求める。

$$Rel(b) = \sum_{i=0}^n Sta(U_s) \quad (3)$$

ここで、 s は主題語を示し、 n は主題語と兄弟語の共通する上位概念の総数を、 U_s は上位概念語を、 $Sta(U_s)$ は U_s の重みを示す。

東京の兄弟語について式(3)を用いて計算した結果の上位10件を表2に示す。ここで兄弟語の重みの上位10件を対立語の候補として取得する。

表2 東京における兄弟語の重みの計算結果

兄弟語 b	兄弟語の重み $Rel(b)$
ニューヨーク	35.1
シドニー	34.2
ロンドン	33.9
パリ	33.9
居酒屋	31.8
バンクーバー	30.7
北京	30.7
ひまわり	30.6
上海	26.3
大阪	24.5

以上より取得した対立語の候補から対立語を抽出するために語の認知度を求める。我々の提案する語の認知度は、検索結果数が近い語ほど認知度が近いと考え、ある語をクエリとした Google 検索結果数をその語の認知度とする。つまりは、主題語と対立語の候補それぞれの検

索結果数を各々の語の認知度として取得する。認知度を取得後、式4に示すように、主題語と対立語の候補の認知度の比率を求め、その比率が大きい対立語の候補上位3件を対立語とする。

$$Con(s, c) = 1 - \frac{\log\{|Cog(s) - Cog(c)|\}}{\max\{Cog(s), Cog(c)\}} \quad (4)$$

ここで、 s は主題語を示し、 c は対立語候補を示す。また $Cog(s)$ は主題語 s の検索結果数、 $Cog(c)$ は対立語候補 c の検索結果数を示す。

東京の対立語の候補となる語10件における認知度の対比率の結果を表3に示す。この結果より主題語「東京」に対して対立語「大阪」がとれていることがわかる。

表3 東京における対立語候補の検索結果数及び対比率

対立語候補 c	検索結果数 $Cog(c)$	対比率 $Con(s, c)$
ニューヨーク	16100000	0.05
シドニー	2750000	0.01
ロンドン	16200000	0.05
パリ	17600000	0.05
居酒屋	40700000	0.17
バンクーバー	2430000	0.01
北京	62500000	0.52
ひまわり	9050000	0.03
上海	130000000	0.41
大阪	470000000	0.69

3.4 対立記事抽出のための記事アスペクトと対立語のペア決定

対立語記事を抽出するために、3.2節と3.3節に提案したように閲覧記事から取得した記事アスペクトと対立語をクエリとして対立記事の候補となる記事を取得する。しかしながら一つの記事アスペクトに対して複数の対立語があるため、記事アスペクトと対立語の複数のペアが存在する。その中で、閲覧記事に対するある記事アスペクトによる対立記事を抽出するのに適した記事アスペクトと対立語の記事を決定する必要がある。そこで、我々は対立語と記事アスペクトがよく共起している場合、2つの語の関係が深いと考え、対立記事を抽出する為の記事アスペクトと対立語のペアを決定する。具体的には、3.2節で取得した記事アスペクトと3.3節で取得した対立語の共起度を求め、その共起度が最も高い対立語をその記事アスペクトの対立語とし、記事アスペクトと対立語のペアを決定する。ここで、記事アスペクトは複数あり、各々の記事アスペクトに対して対立語は一つとし、記事アスペクトの個数分のペアを決定する。ここで、共起度を求めるには、Dice係数、cos類似度、Jaccard係数、Simpson係数など様々な尺度がある[8]が、本研究では式(5)に示すようにDice係数を用いる。

$$D_{ar} = \frac{2 \times RA_{ar}}{Cog(a) + Cog(r)} \quad (5)$$

ここで、 a は記事アスペクトを r は対立語を示し、 RA_{ar} は記事アスペクト a と対立語 r で組み合わせた時の検索結果数、 $Cog(a)$ は記事アスペクト a の検索結果数、 $Cog(r)$ は対立語 r の検索結果数、 D_{ar} は記事アスペクト a と対立語 r のDice係数の結果を示す。

ここで、主題語が「アサヒビール」の記事に対して抽出した記事アスペクト「ビール」、「販売」、「クリア」、「両立」、「度数」の検索結果数を表4に示す。

表4 閲覧記事の記事アスペクトの検索結果数

記事アスペクト	ビール	発売	クリア	両立	度数
検索結果数	20700	47200	3260	497	204

また、主題語から抽出した対立語「サッポロビール」、「住友化学」、「関西電力」の検索結果数を表5に示す。

表5 対立語の検索結果数

対立語	サッポロビール	住友化学	関西電力
検索結果数	238	66	3020

さらに、閲覧記事の記事アスペクトと対立語におけるそれぞれの組み合わせの検索結果数を表6に示す。縦は対立語、横は記事アスペクトを示している。

表6 各組み合わせの検索結果数

	ビール	発売	クリア	両立	度数
サッポロビール	242	178	7	5	4
住友化学	2	12	0	1	0
関西電力	182	200	67	17	0

閲覧記事の記事アスペクトの検索結果数及び対立語の検索結果数、閲覧記事の記事アスペクトと対立語におけるそれぞれの組み合わせの検索結果数を用いたDice係数の計算結果を表7に示す。

表4より最も検索結果数が多い記事アスペクトは「発売」であり、表5より最も検索結果数の多い対立語は「関西電力」であるのに対し、表6より最も検索結果数の多い記事アスペクトと対立語を組み合わせは「ビール」「サッポロビール」であることが分かる。また、表7のDice係数の計算結果より記事アスペクト「ビール」には対立語「サッポロビール」、「販売」には「関西電力」などのペアが決まる。

3.5 対立記事の抽出

3.4節により決定したある記事アスペクトと対立語のペアをクエリとして対立記事の候補を取得する。閲覧記事と類似する対立記事の候補が対立記事となると考える。しかしながら対立記事は内容が閲覧記事と類似しているとは限らないため、単純に記事の内容で類似度計算を行っては対立記事となる記事の類似度が高いとは限らない。そこで、我々は記事の記事アスペクトに着目し、

表 7 Dice 係数の計算結果

	ビール	発売	クリア	両立	度数
サッポロビール	0.0231	0.0075	0.0040	0.0136	0.0183
住友化学	0.0002	0.0005	0	0.0036	0
関西電力	0.0153	0.0080	0.0213	0.0097	0

閲覧記事と記事アスペクトが類似している記事が対立記事となると考え、対立記事候補から記事アスペクトを抽出し、閲覧記事の記事アスペクトとの類似度を求め、最も高い類似度を持つ記事を対立記事とし、提示する。この時、類似度計算には \cos 類似度を用いる。

3.6 ユーザインタフェース

対立記事を提示するためのユーザインタフェースを図 2 に示す。

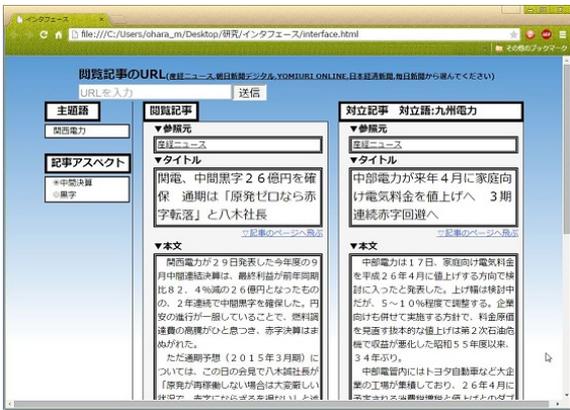


図 2 ユーザインタフェース

ユーザが閲覧記事の URL を入力することで、対立記事の提示を行う。その際、閲覧記事から抽出した主題語を表示する。さらに記事アスペクト毎に対立記事は異なるため、抽出した記事アスペクトの一覧を表示し、任意の記事アスペクトを選択することで対応する対立語と対立記事を表示する。

4 評価実験

提案手法の有用性を計る為に、評価実験を行った。実験データは閲覧記事を任意の記事 20 件とし、被験者は 20 代男女 8 名とする。被験者はシステムを使用して、主題語の抽出結果、記事アスペクトの抽出結果、そして対立記事の抽出結果について各々 5 段階評価を行った。5 段階評価では、5 が最も類似しているであり、1 が最も類似していないとした。

主題語に関する評価値の割合を図 3 に示す。

図 3 より評価値 4~5 の割合が 95% となり、閲覧記事から主題語が適切に抽出できていると言える。主題語の評価が低かった例としては「BSE 発生、ノルウェー産牛肉輸入停止 厚労省発表」に関する記事で、主題語は

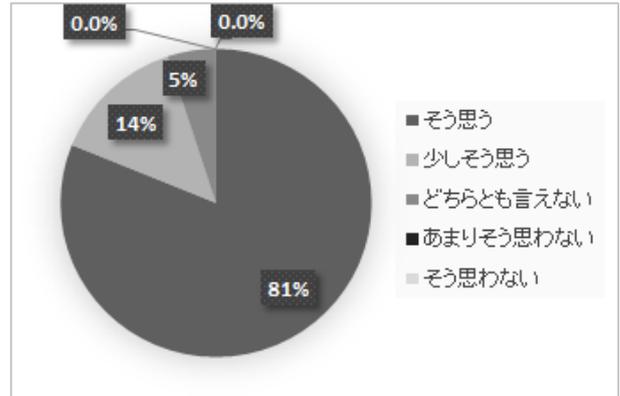


図 3 主題語における評価値の割合

「ノルウェー」が抽出された。評価の低かった原因としては、適切な主題語が「ノルウェー」だけでなく「BSE」や「牛肉」、「厚生労働省」など複数考えられる点や、そのうち「牛肉」という語が一般名詞である点が挙げられる。そのため、主題語を抽出する際は一般名詞も考慮する必要があると考えられる。

閲覧記事の記事アスペクトに関する評価値の割合を図 4 に示す。

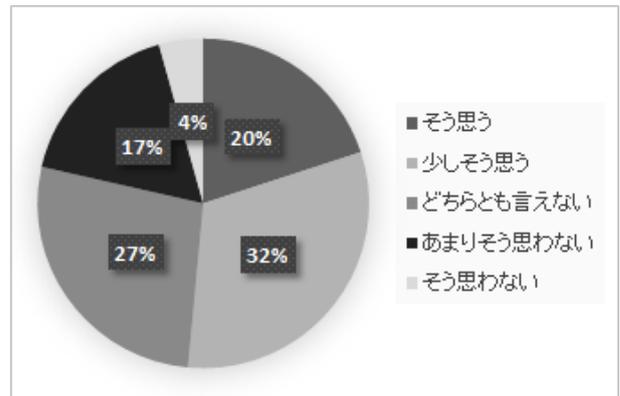


図 4 閲覧記事の記事アスペクトにおける評価値の割合

図 4 より評価値 4~5 の割合が 52% となった。記事アスペクトの評価の高かった例としては「鹿児島県知事、川内原発再稼働に同意を表明」に関する記事で、記事アスペクトが「稼働」や「同意」、「知事」、「原発」となり、記事アスペクトのみで記事の概要を把握できるような語が抽出されたためだと考えられる。評価の低かった例としては、「KDDI が春モデル発表 ボルテ対応スマホを 5 機種投入 アンドロイド搭載のガラケーも」に関する記事で、記事アスペクトが「機種」や「対応」、「発売」、「モデル」、「向け」、「搭載」となり、特に「対応」、「発売」、「向け」といった語の評価が低かった。原因として、評価の低かった記事アスペクトはこの記事で抽出された主題語である「スマホ」特有の語ではなく、様々な語に対して用いられやすい語、つまり汎用性の高い語である

ことが分かった。そのため、汎用性の高い語には重みを小さくすることや、前後の語と組み合わせることで「向け」ではなく「子供向け」など汎用性を低くする必要があると考えられる。

対立記事に関する評価値の割合を図5示す。

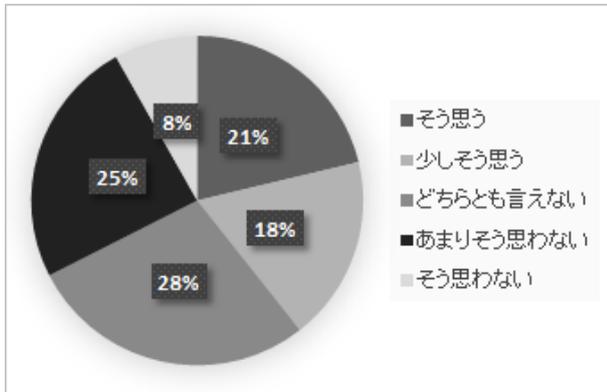


図5 対立記事における評価値の割合

図5より評価値4~5の割合が39%という低い結果となり、適切に抽出できているとは言えなかった。対立記事の評価の高かった例としては、「VIP 会見にイチロー『ただただ恐縮している』」に関する記事に対して、対立記事が「『平成の怪物』松坂、悩んだ末の日本復帰…『世界一目指す』に魅力」に関する記事だった。評価の高かった原因としては、「会見」や「入団」、「球団」といった記事アスペクトが類似度計算に大きく影響したと考えられる。また、評価の低かった例としては、「ドイツ、1月の物価0・3%下落 5年4カ月ぶりにマイナス」に関する記事に対して、対立記事が「【ワイン】フランスがトップの供給国、次いでイタリア、チリ!」に関する記事となった。評価の低かった原因として、抽出した対立記事は「消費」という記事アスペクトのみが類似度計算に影響しており、対立記事の候補となる記事から記事アスペクトを抽出する際に閾値0.6以下の記事アスペクトとならなかった語を類似度計算で考慮していなかった点や、実験の結果、記事アスペクトである「消費」の評価が低かった点、さらに対立語である「フランス」に関する記事が少なく、適切な対立記事が抽出できていなかった点分かった。そのため、記事アスペクトの適切な抽出や、類似度計算の結果についても閾値を設定することで別の組み合わせの対立語と記事アスペクトで対立記事を抽出する必要があると考えられる。

5 まとめと今後の課題

ニュース記事の重要性の理解を支援することを目的とし、ユーザの閲覧している記事に対する記事アスペクト毎の対立記事を抽出・提示する手法を提案した。具体的には、まず閲覧記事から主題語と記事アスペクトを抽出

し、Wikipediaの階層構造を用いて主題語に関する対立語を抽出した。次に、記事アスペクトと対立語の共起関係から、記事アスペクトと対立語の組み合わせを決定し、その組み合わせをクエリとして対立語の候補となる記事を取得した。取得した記事についてそれぞれ記事アスペクトを取得し、閲覧記事の記事アスペクトとの類似度を求めることで、類似度の最も高い記事を対立記事として提示を行った。

今後の課題として、評価実験の結果より閲覧記事の記事アスペクトの抽出と対立記事の抽出が適切でなかったことが分かったため、記事アスペクトに関しては、語の汎用性を考慮し、前後の単語と組み合わせることで適切な記事アスペクトの抽出を行う。対立記事に関しては、類似度計算に問題があると考え、さらなる手法を提案し、精度の向上を行う。また、ニュース記事を取得する対象としているサイト数を増やし、対立記事抽出の制度の向上と閲覧記事の多様性を図る。さらに、対立語を抽出する際、主題語と共通する上位概念に対して記事の内容を考慮することで、より主題語と関係の強い対立語の抽出を行う。また、ユーザインタフェースの利便性の評価実験を行う。

参考文献

- [1] 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”, 電子情報通信学会技術研究報告 DE, データ工学, 107(131), pp.277-282, 2007.
- [2] 池田大介, 藤木稔明, 奥村学, “blog とニュース記事の自動対応付け”, 言語処理学会第11回年次大会論文集, pp.1030-1033, 2005.
- [3] 田中祥太郎, 田中克己, “ニュース記事の理解支援のための背景知識抽出と補完”, 情報処理学会研究報告, データベース・システム研究会報告, pp.1-6, 2014.
- [4] 佐藤吉秀, 川島晴美, 佐々木努, 奥雅博, “時系列ニュース記事における最新話題語抽出方法”, 電子情報通信学会技術研究報告 NLC, 105(203), pp.1-6, 2005.
- [5] 大島裕明, 田中克己, “両方向構文パターンを用いた Web 検索エンジンからの高速関連語発見手法”, 日本データベース学会 Letters, 7(3), 2008.
- [6] 真下遼, 灘本明代, “対立語抽出に基づく Web ニュースからの漫才ロボット台本自動生成手法の提案”, DEIM Forum 2014 C2-4, 8pages, 2014
- [7] 佃洗根, 大島裕明, 山本光穂, “語の認知度と同位語間の関係に基づく意外な情報の発見”, 日本データベース学会論文誌 DBSJ journal, 11(3), pp.21-26, 2013.
- [8] Manning, Christopher D, Schtze, Hinrich, “Foundations of statistical natural language processing”, MIT press, 1999.
- [9] 打田裕樹, 吉川大弘, 古橋武, 平尾英司, 井口浩人, “Web ユーザレビュー解析のための文章データマッピング手法の提案”, 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, 24(0), pp.168-168, 2008.
- [10] 宮部泰成, 高村大也, 奥村学, “異なる文書中の文間関係の特定”, 情報処理学会研究報告 自然言語処理研究会, 105(203), pp.35-42, 2005.