ステークホルダーマイニングとそれに基づくニュース報道の比較

小川　達也† 　馬　　強†† 　吉川　正俊††

† 京都大学工学部情報学科 〒606-851 京都府京都市左京区吉田本町
†† 京都大学大学大学院情報学研究科 〒606-851 京都府京都市左京区吉田本町
E-mail: †ogawa@db.soc.i.kyoto-u.ac.jp, ††{qiang,yoshikawa}@i.kyoto-u.ac.jp

あらまし　我々はニュース報道の偏向は利害関係に起因することが多いことに着目し，利害関係者（ステークホルダー）に関する記述と評価を比較して，報道の偏向の分析を支援する手法を提案する．事象における利害関係を記事から抽出するため，エンティティの関係を表現する語彙の辞書を試作する．この関係辞書を用いて文の構文構造を解析し，ステークホルダーの関係グラフを生成してステークホルダーを抽出する．さらに，応用として発信者ごとに報道されているステークホルダーを比較してニュースにおける偏向を明らかにする．

キーワード　ニュース報道の比較 ステークホルダー 信頼性 Web マイニング Web コンテツ解析

Stakeholder Mining and Its Application for News Comparison

Tatsuya OGAWA†, Qiang MA††, and Masatoshi YOSHIKAWA††

† School of Informatics and Mathematical Science
Faculty of Engineering, Kyoto University
†† Graduate School of Informatics
Kyoto University
E-mail: †ogawa@db.soc.i.kyoto-u.ac.jp, ††{qiang,yoshikawa}@i.kyoto-u.ac.jp

Abstract　In this paper, we propose a novel method for bias analysis of news articles by comparing descriptions about stakeholders based on the presumption that interests often induce media bias. We construct a dictionary that provides scores ranging from 0.0 to 1.0 to indicate states of relationship between entities. An analysis of sentence structure with the dictionary enables us to make relationship graph of entities and we extract stakeholders by using the graph. Furthermore, as an application, we elucidate media bias by comparing descriptions of stakeholders. We also show some experimental results.

Key words　Comparing news articles Stakeholder Credibility Web mining Web contents analysis

1. 緒　　論

ニュース記事には，書き手やスポンサーの意図などによる偏りがある．そのため，ニュース記事の読者は事象の理解に偏りが生じる可能性がある．ニュースの理解を支援するため，関連記事を収集して比較分析を行う研究が多くあるが，差異の発見はユーザに委ね，提示手法に依存する手法（[1]〜[3]）が多く，差異や偏りを分析するための明確なモデルや基準がまだ不十分である．

偏りは利害関係から生じることが多いため，記事に記述されている人物，組織などの実体とその利害関係から記事の偏向を発見することが可能である．本研究では，我々は利害関係者（ステークホルダー）に着目し，人物や利害関係に関する記事を分析して，ニュース報道における偏向の分析を行う手法を提案する．

本稿ではステークホルダーを「ある事象に参加しており，他の参加者とその事象に関して利害関係を持つもの」として定義する．また，利害関係を持つ可能性のある実体，例えば国，人，組織などをエンティティとする．

ステークホルダーは他のステークホルダーと利害関係を持つため，利害関係を持つエンティティはステークホルダーと見なせる．よって，記事から事象における利害関係を分析することでステークホルダーを抽出できる．

エンティティ間の利害関係の特定は文の構造を用いて行う．そこで我々はエンティティ間の関係を分析するために，関係の分析に適した関係構文構造を定義する．また，利害関係は単語によって表現されるため，エンティティの関係を表現する語彙の辞書である関係語辞書（RelationshipSentiWordNet）を構築

した．この辞書は各語彙にポジティブ，ネガティブ，オブジェクティブの 3 つの数値を割り当てており，ポジティブの数値が高い程エンティティ間の関係が良いことを表現し，ネガティブの数値が高い程悪いことを表す．オブジェクティブは客観的な程度を意味する．関係語辞書に含まれる語彙と関係構文構造を用いて記事から抽出した利害関係に基づいてグラフを構築する．このグラフから利害一致及び利害対立している関係をまとめることによって事象におけるステークホルダーを発見できる．そして，記事から抽出したステークホルダーと関係構文構造を用いて，記事のステークホルダーに対する評価を求める．評価値は WordNet [4] から作成した感情辞書である SentiWordNet [5] を用いて分析を行う．

さらに，ステークホルダーマイニングを用いた，ニュース報道の偏向分析手法を提案する．ステークホルダーマイニングによって事象に対するニュース記事の視点を定量化することができ，ニュース記事の偏向分析が可能となる．

以下，2 章では関連研究ついて述べる．3 章ではニュース記事から利害関係及びステークホルダーへの評価を抽出するステークホルダーマイニングについて説明する．4 章ではでステークホルダーマイニングに基づいたニュース報道の比較を行う．5 章では評価実験について記す．

2. 関 連 研 究

ステークホルダーの観点を用いた比較の研究として，Liu [1] らによるものがある．これはあるニュース記事に対してその文書に関係する国をステークホルダーとし，それらの国の同じ事象を扱う記事を集め，国毎に数個表示し，その国家の利害を代表する意見をハイライトしてユーザに提示するシステムを作成している．これによりその記事の国の視点の意見を見ることができる．

NewsCube [2] は事象の様々な観点を提示するシステムである．ニュース記事から事象の様々な側面を抽出して提供することによって，ユーザにより深いニュースの理解や，偏った認識を防ぐことを促す．

The Comparative Web Browser(CWB) [3] は，ユーザが読んでいる記事の文章と類似する文章を持つ記事を検索し，提示する．このシステムにより，ユーザは記事を比較しながら読むことができる．

TVBanc [6] は，トピックと記事の視点に基づいてニュースを比較し，ニュースの偏りと多様性を分析するシステムである．TVBanc は，まずコンテンツの構造からトピックと視点を抽出し，関連するニュースを様々なメディアから収集する．そして，収集したニュースを分類することによって偏りと多様性を分析する．

石田 [7] らの研究では，報道元ごとの特徴をユーザに提示するシステムを提案している．あるエンティティに関する記述の SVO 構造を分析することで，報道元ごとに特徴的な記述を抽出している．

ニュース記事を提供するメディアの偏りについての分析は Grefenstette [8] らの研究がある．これは検索対象をニュースに限定しているニッチブラウザである Google News [9] を用いて，あるエンティティに関する記事を収集し，そのエンティティの名前が記述されている周囲の文章の単語からそのエンティティに対する評価の分析を行う．

これらの研究に対し本研究では，記事からステークホルダーとその利害関係の抽出を行う．そして，ステークホルダーを用いて事象に対する記事の視点を分析し，偏りを定量化することによって偏向分析の支援を行う．

3. ステークホルダーマイニング

ステークホルダーマイニングではニュース記事を入力とし，記事からステークホルダーと利害関係及びステークホルダーに対する記事の記述極性を分析し，分析結果のグラフを出力する．

ステークホルダーマイニングは次の手順で行う．

（1） ステークホルダーの候補の抽出

（2） 利害関係の分析と関係グラフの作成

（3） 関係グラフによるステークホルダーの抽出

以下に詳細を記す．

3.1 ステークホルダーの候補抽出

まず与えられた記事から文の集合を生成する．ステークホルダーの定義より，ステークホルダーの候補となるのは国，人，組織が考えられる．そこで，各文に対して固有表現抽出を行う．ステークホルダーの候補として地名，人名，組織名の 3 つの固有表現タイプを抽出する．

3.2 利害関係の分析に基づくステークホルダー抽出

ステークホルダーは利害関係を持つため，ステークホルダーの関係について言及する文には利害関係を表す語が含まれる．よって，文に 2 つ以上のステークホルダーの候補が存在し，かつステークホルダーの候補間の利害関係を表現する語が文中に含まれているならば，ステークホルダーを発見することが可能である．

3.2.1 関係語辞書

文中の利害関係を表現する語を特定するために，WordNet を用いて関係語辞書（RelationshipWordNet）を作成する．WordNet では全ての単語を，同じ意味を持つまとまり (synset) に分類し，各 synset は他の synset と異なる概念を表す．関係語辞書は WordNet の全ての synset に対し，その synset が持つ意味が利害一致を表現する場合はポジティブ，利害対立を表す場合にはネガティブの数値を割り当てる．synset に割り当てられたポジティブの数値が高い程その synset に含まれる語彙は利害一致の関係が強いことを表し，ネガティブの数値が大きい程その synset に含まれる語彙は利害対立の関係が強いことを表現する．

我々は SentiWordNet [5] の作成手法を応用し，関係語辞書の作成を行う．以下に手順を示す．

（1） まず手動でオブジェクティブ，ポジティブ，ネガティブの 3 つの種セット L_o，L_p，L_n を作成する．The Longman Defining Vocabulary の 2193 語から synset を作成し，synset が含む単語が表現する利害関係に基づいてポジティブ, ネガティブ, オブジェクティブのラベルを割り当てることによって種セッ

トを作成した．ある語句が，ステークホルダー間の記述に用いられた場合，その語句が表現する関係の良し悪しに適したラベルを割り当てる．種セットの例を表 1 に記す．

表 1　種セットの例

オブジェクティブ	ポジティブ	ネガティブ
say	comfort	attack
go	depend	blame
act	favorite	complain

（2） 次に，L_p, L_n を K 回拡張することによって訓練セット T_o^K，T_p^K，T_n^K を得る．ただし，任意の K に対し $T_o^K = L_o$ である．拡張は，WordNet で定義される sysnet 間の関係を用いる．T_p^{K-1}（または T_n^{K-1}）に含まれる synset と関係を持つ sysnet を，T_p^{K-1}（T_n^{K-1}）の訓練セットに追加することにより T_p^K（T_n^K）を得る．拡張に用いる synset 間の関係は， *directantonomy similarity*, *derived-from*, *partains-to*, *attribute*, *also-see* である．これら七つの関係のうち，*directantonomy* 以外の関係はポジティブまたはネガティブが一致する関係であるため，ポジティブ，ネガティブのラベルを保持して T_p^{K-1}（T_n^{K-1}）の訓練セットに追加することにより T_p^K（T_n^K）を得る．*directantonomy* は互いに対立する意味を持つ synset 間の関係であるため，ポジティブ，ネガティブを入替えて T_p^{K-1}（T_n^{K-1}）の訓練セットに追加する．

（3） 全ての synset に対して，その synset が持つ意味の文章をコサイン正規化 *tf·idf* を用いてベクトル化する．$K = 0, 2, 4, 6$ の訓練セットを二つの学習機械（Rocchio 及び SVMs(注1)）でそれぞれ学習を行うことで，WordNet の全ての synset をオブジェクティブ，ポジティブ，ネガティブの 3 クラスに分類を行う分類器を 8 つ作成する．

（4） この八つの分類器のオブジェクティブ，ポジティブ，ネガティブの結果の割合に応じて synset にオブジェクティブ，ポジティブ，ネガティブの数値を割り当て，関係語辞書の構築を行う．例えば，ある synset に対する分類器による結果のうち，二つがオブジェクティブ，五つがポジティブ，一つがネガティブとする．このとき，その synset に割当てる数値は，オブジェクティブが 0.250($\frac{2}{8}$)，ポジティブが 0.625($\frac{5}{8}$)，ネガティブが 0.125($\frac{1}{8}$) となる．

関係語辞書の synset のうち，ポジティブ，ネガティブのいづれかの数値が 0 でない synset に含まれる単語を関係語と呼ぶ．関係語辞書と SentiWordNet を組み合わせて用いることにより記事の分析を行う．

3.2.2 関係構文構造

ステークホルダー間の関係は文の構造と単語で表されるため，文の構造による関係の解析のために関係構文構造を定義し，関係語辞書を利用して利害関係の抽出を行う．

Stanford Dependencies [10] [11] は文における単語間の依存関係を以下の形式で表現する．

(注1)：学習機械はAndrew McCallum ' s Bow package (http://www-2.cs.cmu.edu/m̃ccallum/bow/)，Thorsten Joachims' SVM^{light} (http://svmlight.joachims.org/) を用いた．

表 2　依存関係の操作

依存関係	操作
conj	この関係を関係構文構造から削除し，*governor* の親を *dependent* の親にする．さらに，*governor* と *dependent* が共に動詞ならば，*dependent* 以外の *governor* の子を *dependent* の子にする
appos	conj の場合と同じ
rcmod	*governor* と *dependent* を入れ替える
cop	rcmod の場合と同じ

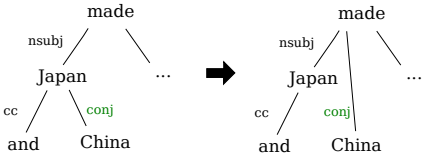

図 1　conj の操作

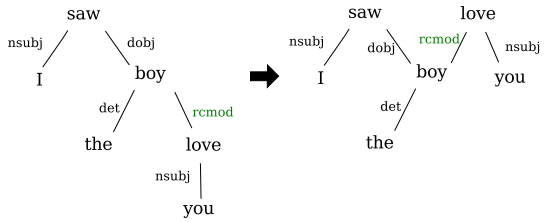

図 2　rcmod の操作

type (*governor*, *dependent*)

type は単語 *governor* と単語 *dependent* の間の依存関係の型である．Stanford Dependencies によって得られる依存関係を，*governor* を親に，*dependent* を子にすることによって木構造を作成する．

ステークホルダー間の関係を特定するためには，関係語と動詞とエンティティの構文構造上の位置関係が重要であり，それを木構造の先祖，子孫関係に基づいて判定するため，まず，木構造の操作が必要である．そこで，木構造に表 2 の依存関係に対応する操作を行うことによって関係構文構造を得る．

表 2 の依存関係のうち，“conj” は “and” で結ばれた 2 つの単語の関係を表す．“appos” は単語が “,” で結ばれた 2 つの単語の関係を表わす．“rcmod” はある名詞が関係代名詞などで修飾されている場合，その名詞と関係代名詞節の動詞との関係を表す．“cop” は be 動詞とその be 動詞を補完する単語の関係を表す．例えば，“Tom is an honest man.” という文では，“is” と “man” の関係である．図 1 は “Japan and China made ...” の文の例であり，“conj(Japan, China)” の依存関係を含む関係である．この場合は “conj(Japan, China)” の関係を削除し，“Japan” の親を “China” の親とする．図 2 は “I saw the boy you love.” という文の例であり，“rcmod(boy, love)” の関係を削除し，“love” を “boy” の親にする．

次の文の関係構文構造は図 3 のようになる．

> Kerry said that WTO, located in Geneva, and rapidly growing China talked with Japan which concluded an good alliance with India.

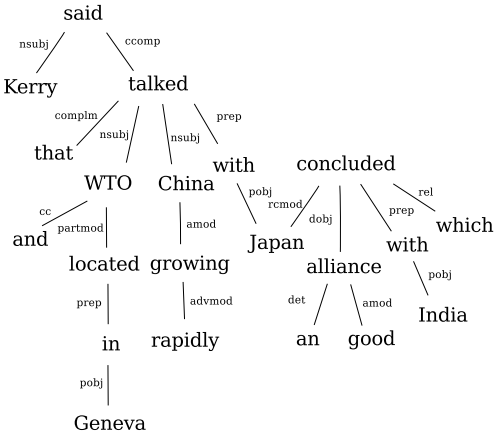

図 3　関係構文構造の例

3.2.3　利害関係の抽出

関係構文構造を利用して，文に記述されているステークホルダー間の関係を求める．利害関係は以下の形式で表現する．

$$(\{S_1, S_2\}, P, N)$$

ここで，S_1 と S_2 はステークホルダーを表し，P と N はそれぞれ S_1 と S_2 の関係の利害一致の程度，利害対立の程度を表す．P と N の両方の値が 0 でない値を持つことは，事象においてある部分では利害一致し，他の部分では利害対立することを表す．

文中のそれぞれの関係語 w_R に対して，以下の処理を行う．ただしこれ以後，二つの節点が “最も近い” とは，その二つの節点間のパス上にある節点の数が最も少いことと同義であるとする．関係語 w_R と w_R の祖先にあるそれぞれのパスにおいて最も w_R に近い動詞の集合を $V(w_R)$ とする．また，動詞 v の子孫のそれぞれのパスの最も近いエンティティの集合を $E(v)$ とする．

手順 1　w_R から $V(w_R)$ までのパスにエンティティがあれば，w_R はそのエンティティの修飾語と考えられる．よって w_R はこの利害関係を表現しないため，次の関係語へ．

手順 2　w_R から $V(w_R)$ までのパスにエンティティがなければ，v を根とする関係構文構造の部分木について，v を V としたときの表 3 に記す構文タイプに応じて利害関係を抽出する．v が表 3 に記す構文タイプのいづれでもなければ $V(w_R)$ は利害関係を表現しないため，次の関係語の処理を行うため，手順 1 へ．

手順 3　v が表 3 に記す構文タイプのいづれかであれば，構文タイプに従って利害関係を抽出する．得られた利害関係が初めての抽出した関係であるならば，P と N の数値の初期値を共に 0 とする．

手順 4　関係語辞書から，w_R の持つポジティブとネガティブの数値を得る．抽出された利害関係の P と N の数値にそれぞれ加算する．このとき，否定語が含まれていた場合，つまり単語間の持つ依存関係が ”neg ” のものうち

neg $(governor, v)$, $v \in V(w_R)$

があれば,w_R のポジティブ，ネガティブの数値を逆にして扱う．

neg $(governor, w_R)$

があれば,w_R のポジティブ，ネガティブの数値を逆にする．

表 3　文の構造による利害関係

構文タイプ	利害関係
$N_{es}VN_{eo}$	$\{\{s_s, s_o\} \mid s_s \in N_{es}, s_o \in N_{eo}\}$
$N_{es}VN$	$\{\{s_{s1}, s_{s2}\} \mid s_{s1} \neq s_{s2}, s_{s1} \in N_{es}, s_{s2} \in N_{es}\}$
$N_{es}VN_{\phi}$	$\{\{s_{s1}, s_{s2}\} \mid s_{s1} \neq s_{s2}, s_{s1} \in N_{es}, s_{s2} \in N_{es}\}$
NVN_{eo}	$\{\{s_{s1}, s_{s2}\} \mid s_{s1} \neq s_{s2}, s_{s1} \in N_{eo}, s_{s2} \in N_{eo}\}$

表 3 の構文タイプについて，N_{es} は $E(v)$ のうち v の主部にあるものが存在すること，N_{eo} は $E(v)$ のうち v の述部にあるものが存在することを表す．N は，主部にある場合 $N_{es} = \phi$ を満たしかつ主部に名詞が存在することを表し，述部にある場合 $N_{eo} = \phi$ を満たしかつ述部に名詞が存在することを表す．N_{ϕ} は V の述部に名詞が存在しないことを表現する．また V は動詞である．

以上で利害関係を持つエンティティが特定できたのでステークホルダーが得られる．

利害関係の抽出の例として図 3 の文を挙る．この文には “talked” と “alliance” という 2 つの関係語がある．“talked” を w_R とすると $V(w_R)$ は {talked} であるので，“talked” の子孫の全てのパスのうちそれぞれ最も talked に近いエンティティは，WTO，China，Japan の三つである．“talked” が持つ文の構造タイプは $N_{es}VN_{eo}$ であるので，{WTO，Japan} と {China，Japan} の関係が抽出できる．これらの関係は “talked” のポジティブ，ネガティブの数値を持つ．関係語辞書より，“talked” はポジティブが 0.250，ネガティブが 0.00 であるとする．{WTO，Japan} と {China，Japan} の利害関係は，“talked” 以外の関係語によって抽出されないので，この文においては

$$(\{WTO, Japan\}, 0.227, 0.00)$$

$$(\{China, Japan\}, 0.227, 0.00)$$

となる．“alliance” は同様に {Japan，India} の関係を表す．“alliance” はポジティブが 0.750，ネガティブが 0.00 であるとすると，{Japan，India} の利害関係は以下のようになる．

$$(\{Japan, India\}, 0.750, 0.00)$$

3.2.4　ステークホルダーに対する記述の極性分析

記述極性とはステークホルダーに対する記述の特徴である．記述極性は，ステークホルダーに関する記述の特徴をポジティブとネガティブの数値によって表す．この記述極性によって，事象に対する記事の視点を定量的に扱うことが可能となる．

抽出したステークホルダーに対する記述の極性分析には，関係構文構造と感情語辞書を用いる．関係構文構造によって，ステークホルダーに対する記述を特定する．そして，感情語辞書を用いて記述の極性の分析を行う．

WordNet を用いた感情辞書の関連研究には SentiWordNet がある．SentiWordNet は WordNet のそれぞれの synset に，*Obj(s)*，*Pos(s)*，*Neg(s)* の 3 つの数値を合計が 1 となるように割り当てられている．3 つの数値はその synset に含まれている単語がどれだけオブジェクティブ，ポジティブ，ネガティブかを表す．

SentiWordNet のポジティブ，ネガティブのいづれかの数値が 0 でない単語を感情語と呼ぶ．記述極性の抽出は，文中の各感情語 w_S に対し，以下の処理を行う．感情語 w_S と，その感情語 w_S の先祖とを結ぶそれぞれのパス上において，最も w_S に近い動詞の集合を $V(w_S)$，動詞 v の子孫のそれぞれのパスの最も近いエンティティの集合を $E(v)$ とする．

手順 1　w_S から $V(w_S)$ のパス上にエンティティがある場合，w_S はそのエンティティを修飾していると考えられるので，そのエンティティへの評価に w_S の P，N の数値を加算する．

手順 2　w_S から $V(w_S)$ のパス上にエンティティがない場合，w_S は $E(v)$（ただし $v \in V(w_S)$）のそれぞれの要素であるエンティティへの評価とする．よってそれらのエンティティへの評価に w_S の P，N の数値を加算する．

図 3 の文においては，感情語は “growing” 及び “good” である．“growing” を w_S とすると w_S から $V(w_S)$ のパス上にエンティティである “China” がある．よって，“growing” は “China” への修飾であるので，“growing” のポジティブとネガティブの数値を “China” の記述極性の値に加算する．“good” を w_S とすると w_S から $V(w_S)$ のパス上にエンティティは存在しない．よって，“good” は $E(v)$（ただし $v \in V(w_S)$），つまり日本とインドについての肯定的な記述極性である．よって “good” のポジティブとネガティブの数値を，日本とインドにそれぞれ加算する．これは “good” によって日本とインドが良い同盟を結んだと記述していることを意味する．

3.3　関係グラフの構築

3.3.1　関係グラフの定義

ニュース記事は，一般にその文書に含まれるそれぞれの文が事象の断片を表現し，文書全体で事象を表現する．よって，まず一文ごとに利害関係を持つステークホルダーと関係の数値化を行う．そして，全ての文の関係を利害関係ごとにポジティブとネガティブの数値をそれぞれ加算することによって，文書全体が表現する利害関係を求める．記述極性も同様に，同一のステークホルダーに対する全てのポジティブとネガティブの評価をそれぞれ加算することによって，記事全体の特徴を求める．

利害関係はステークホルダー間の関係であるため，節点をステークホルダーとし，枝を利害関係とするグラフ構造とみなすことができる．これを関係グラフと名付ける．

[定義 1]（関係グラフ）　関係グラフ G を以下に定義する．

ステークホルダーを節点 v，利害関係を枝 e とする．また，節点集合を V，枝集合を E，$\Re$ を実数集合とする．V，E の元から二組の実数へのそれぞれの写像

$$l : V \mapsto \Re \times \Re$$
$$w : E \mapsto \Re \times \Re$$

とする．ここで，l による節点の像である二組の数値は, その節点が表すステークホルダーに対する記述極性を表す．それぞれポジティブ，ネガティブに対応する．w による枝の像である二組の数値は利害関係の状態を表す．一方は利害一致の程度，他方利害対立の程度である．

g を E の元から V の 2 つの元の集合への写像

表 4　利害関係の種類

利害タイプ	利害状態	説明
5	$P > 0, N = 0$	純粋な利害一致の関係
4	$P > N > 0$	利害一致と利害対立している部分があるが，利害一致の方が強い関係
3	$P = N \neq 0$	利害一致と利害対立している部分があり，それらが同等な関係
2	$N > P > 0$	利害一致と利害対立している部分があるが，利害対立の方が強い関係
1	$N > 0, P = 0$	純粋な利害対立の関係

$$g : e \mapsto \{v_1, v_2\}$$

(ただし，$v_1, v_2 \in V$ である) とするとき，

$$G := \langle V, E, g, l, w \rangle$$

を関係グラフとする．g は，利害関係は二つのステークホルダーの関係であることを意味する．

関係グラフの例を図 4 に挙げる．

3.3.2　関係グラフの集約

最後に，関係グラフから利害関係を利用してステークホルダーをまとめて利害関係の整理を行う．ステークホルダー間の利害関係は，利害関係の P と N の値に応じて表 4 の 5 つの利害タイプに分類できる．

そこで，この 5 つの関係を用いて，利害が一致しているものをまとめることでステークホルダーを導出する．関係グラフの集約アルゴリズムを Algorithm 1 に記す．入出力は共に関係グラフとする．関数 $type(e)$ は表 4 に示す枝 e の利害タイプを返す．また，論理式 1 は次を意味する．利害関係 e を持つステークホルダーを v_1，v_2 とする．$v_1 \in g(e')$，$v_2 \in g(e'')$，$v' = v''$ (だだし，$v_1 \neq v'$,$v' \in g(e')$，$v_2 \neq v''$,$v'' \in g(e'')$) を満す利害関係の集合 E に含まれる，全ての二つの利害関係の組 e'，e'' に対して，$type(e') = type(e'')$ が成り立つならば真となる．論理式 1 を (1) に記す．

$$\begin{aligned}
&\exists e(\\
&\forall e' \forall e'' \forall v_1 \forall v_2 \forall v_1' \forall v_2' \forall v_1'' \forall v_2'' \\
&(\neg((v_2' = v_2'') \wedge (g(e') = \{v_1', v_2'\}) \\
&\wedge (g(e'') = \{v_1'', v_2''\}) \wedge (e', e'' \in E) \\
&\wedge((v_1 = v_1') \wedge (v_2 = v_1'') \wedge (g(e) = \{v_1, v_2\}))) \\
&\vee (type(e') = type(e'')) \\
&)) \qquad (1)
\end{aligned}$$

論理式 2 は次を意味する．利害関係 e を持つステークホルダーを v_1，v_2 とするとき，$v_1 \in g(e')$ が成り立つ全ての $e' \in E$ に対し，$v_2 \in g(e'')$ かつ $v' = v''$ (だだし，$v_1 \neq v'$, $v' \in g(e')$, $v_2 \neq v''$, $v'' \in g(e'', e'' \in E)$ が成り立つ $v_2 \in g(e'')$ が存在しないならば真となる．論理式 2 を (2) に記す．

$$\exists e($$

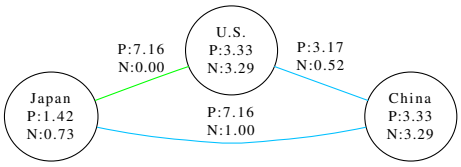

図 4　関係グラフの例

$$\forall v_1, \forall v_2 \forall e' \forall v_1' \forall v_2'$$
$$(\neg((g(e') = \{v_1', v_2'\}) \wedge (e' \in E)$$
$$\wedge((v_1 = v_1') \wedge (g(e) = \{v_1, v_2\}))))$$
$$\vee\neg((\exists e'' \exists v_1'' \exists v_2''(e'' \in E) \wedge (v_2 = v_1'') \wedge (v_2' = v_2'')))$$
$$) \qquad (2)$$

Algorithm 1 のステップ 4 の三つの条件を満す利害関係は，その利害関係を持つ二つのステークホルダーが利害一致の関係である．また，その二つのステークホルダー v_1，v_2 が，他のステークホルダー v と利害関係を持つ場合，v_1 と v_2 は，v とそれぞれ同じ利害タイプの利害関係にある．よって，v_1 と v_2 を一つのステークホルダーと見なしてもよいと考えられる．

f_1 は入力を利害関係 e と利害関係の集合 E の二つとし，利害関係の集合を返す関数である．この関数が返す利害関係の集合は，$g(e) = \{u, v\}$ とするとき，$g(e') = \{u, w\}$ かつ $g(e'') = \{v, w\}$ を満たす利害関係 e', e'' の集合である．これはステークホルダーをまとめる際に無くなる利害関係を意味する．

f_2 は入力を，ステップ 5 においてまとめて得たステークホルダー v，利害関係 e，利害関係の集合 E とする．そして，の利害関係の集合を返す．この関数が返す利害関係の集合は次を意味する．$g(e) = \{u, v\}$ とするとき，$g(e') = \{u, w\}$ かつ $g(e'') = \{v, w\}$ を満たすそれぞれの w と，ステークホルダー v との間の利害関係である．これは，ステークホルダーをまとめる際に新たに生じる利害関係を意味する．

Algorithm 1 関係グラフの集約

Require: G:関係グラフ
Ensure: v:節点
1: $V :=$ 関係グラフ G の節点集合.
2: $E :=$ 関係グラフ G の枝集合.
3: **for all** e in E **do**
4:　**if** $(type(e) = 5) \wedge$ 論理式 1 $\wedge$ 論理式 2 **then**
5:　　$v \leftarrow g(e)$ の二つの節点をまとめて一つの節点を作成.
6:　　$V \leftarrow V \cup v$.
7:　　$E' \leftarrow f_1(e, E)$.
8:　　$E'' \leftarrow f_2(v, e, E)$.
9:　　$E \leftarrow \{E \cup E''\} - \{E' \cup e\}$.
10:　　$V \leftarrow V - g(e)$.
11:　**end if**
12: **end for**

以上の手法により，記事が扱うステークホルダーを発見でき，記事に記述された文章の特徴の発見が可能となる．

4. ステークホルダーを用いたニュース報道の比較

本章では，ステークホルダーマイニングの応用として，ニュース報道の比較による偏向の分析支援について述べる．

まず比較して偏向を分析したい記事に対してステークホルダーマイニングを行う．マイニングの結果として得られたステークホルダーの関係グラフを比較する．ステークホルダーマイニングの分析結果から得られる，1) 記事に記述されているステークホルダー，2) ステークホルダーに対する記述極性，3) 記述されている利害関係，を比較することによって記事の偏りを発見する．

4.1 ステークホルダーの比較

二つの記事を比較するとき，一方の記事の分析結果にしかステークホルダーがない場合がある．これは，一方の記事はそのステークホルダーに対する記述があるが，他方は取り挙げていない可能性があることを意味するため，報道の範囲が偏っていることになる．

例えば 2 つの記事に含まれるステークホルダーが以下の場合，記事 1 はロシアが含まれていないが，記事 2 では含まれている．

- 記事 1 のステークホルダー：日本，北朝鮮
- 記事 2 のステークホルダー：日本，北朝鮮，ロシア

これの 2 つの記事は，一方の記事だけがロシアについて言及している可能性があることを意味する．

4.2 ステークホルダーに対する記述の極性比較

節点の数値 P/N は記事の各ステークホルダーに対する視点を表現している．例えば次の 2 つの記事の場合，記事 1 はハイチに対する記述がネガティブであるのに対し，記事 2 ではハイチに対する記述がポジティブとなる．

- 記事 1：ハイチでは暴動が起こっている．
- 記事 2：ハイチでは救助活動が行われている．

よって P と N の数値から，事象に対する記事の視点が理解できる．

4.3 利害関係の比較

記述されている利害関係からも，記事の特徴や記述範囲の偏りが理解できる．差異として表れるのは，ステークホルダー間の利害関係の有無や，利害関係のポジティブとネガティブの数値の違いがある．

5. 評 価 実 験

提案手法を評価するため, 以下の実験を行った．

（1） 利害関係抽出の評価実験
（2） ステークホルダーに対する極性分析の評価
（3） ステークホルダーマイニングの結果に基づく偏向分析のユーザ評価

また，我々は記事から利害関係を抽出するために作成した関係語辞書の評価を行う必要がある．しかし synset の意味とその synset に割当てられた三つの値から，synset が表現する利害一致，または利害対立の程度を直接評価するのは困難である．そこで我々は，関係語辞書を用いて抽出した利害関係を分析す

表 5　ステークホルダーマイニングの評価結果

	p_{re}	r_{re}	p_{pa}
10 トピックの平均	65.88%	69.12%	67.26%

ることによって，辞書の間接評価を行う．

5.1　実験 1：利害関係の抽出に関する評価実験

データセットとして 10 個のニューストピックを選んだ．それぞれのトピックは，そのトピックに関する五つの英語のニュース記事から構成される．記事はそれぞれのトピックに対して，Google News [9] の関連記事の中から選び，計 50 個事に対して実験を行う．また，各々のトピックに属する五つの記事が記述された日時の差は，7 日以内のものである．

記事に対してステークホルダーマイニングを行うと，その記事に記述されている利害関係が抽出できる．利害関係の抽出に関する評価実験では，データセットの 50 個の記事に対してステークホルダーマイニングを行い，分析対象の 50 個の記事に対し，個々の記事が記述する利害関係をそれぞれのマイニング結果が抽出できているかを評価した．

利害関係の抽出の際に，関係語辞書から単語の数値を取得する必要があるが，文脈に即した意味の語句を抽出するのは困難である．そのため今回の実験では，次の方法をとる．語句 w を含む synset の集合を $S = \{s_1, s_2, \ldots, s_n\}$ とする．S のうち，$P > N$ である synset の集合を S_P,$P < N$ である synset の集合は S_N,$P = N = 0$ である synset の集合を S_O とする．語句 w の P，N，O の数値を以下の式 (3) で与える．

$$Score(w) = m \cdot \Delta \tag{3}$$

ただし，

$$m = max(\frac{|S_P|}{n}, \frac{|S_N|}{n}, \frac{|S_O|}{n})$$

$$\Delta = \begin{cases} ave(S_P) \cdot (1,0,0)^T & (|S_P| > max(|S_N|, |S_O|)) \\ ave(S_N) \cdot (0,1,0)^T & (|S_N| > max(|S_P|, |S_O|)) \\ (0,0,1) & (\text{上記以外}) \end{cases}$$

とする．また，$ave(S_X)$ は，S_X に含まれる各 synset の P，N，O を，それぞれ平均した要素を持つ 1×3 の行列である．

利害関係の抽出では適合率と再現率で評価を行った．それぞれの記事での適合率と再現率をトピックごとに平均し，10 トピックの評価結果の平均を求めた．その結果，適合率 p_{re} は 65.88%，再現率 r_{re} は 69.12% となった (表 5)．

5.2　実験 2：記述の極性分析に関する評価実験

ステークホルダーに対する記述の極性分析の評価実験では，実験 1 と同一の 50 個の記事をデータセットとする．これらのデータセットに対してステークホルダーマイニングを行い，各々の分析結果がステークホルダーに対して記述している文章の極性を抽出できているかを評価する．感情語の数値は取得する感情語を含む全ての synset に対して，ポジティブ，ネガティブの数値のそれぞれの平均とする．

極性分析に関する評価実験ではポジティブ，ネガティブの極性の大小が適合しているか評価を行った．それぞれの記事での適合率をトピックごとに平均し，10 トピックの評価結果の平均を求めた．その結果，適合率 p_{pa} は 67.26% となった (表 5)．

表 6　ユーザ評価の結果

		評点平均	5 及び 4 の割合
質問 1	同国	3.45	40.0%
	異国	3.90	75.0%
	合計	3.67	58.5%
質問 2	同国	3.75	65.0%
	異国	4.10	75.0%
	合計	3.93	70.0%

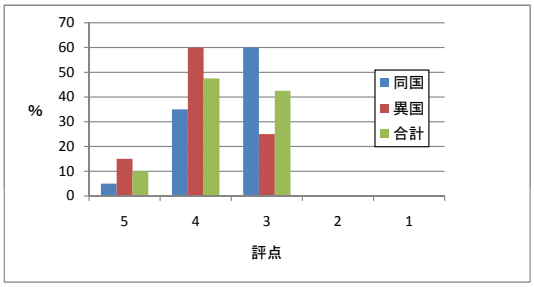

図 5　質問 1 の結果

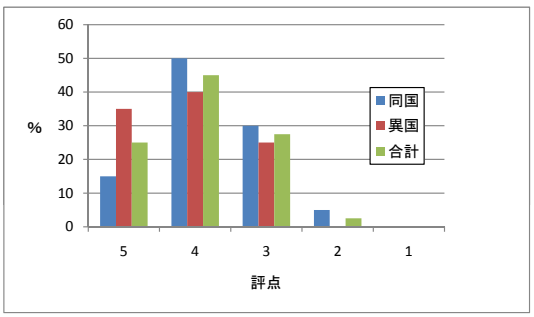

図 6　質問 2 の結果

5.3　実験 3：偏向の分析手法のユーザ評価

ステークホルダーマイニングの分析結果に対してユーザ評価を行う．ユーザ評価によって偏向の分析支援における分析結果の有用性を評価する．

ユーザ評価で対象とするデータセットは実験 1 で用いたデータセットから作成した．実験 1 のデータセットである十個のトピックから，トピック毎に二つの記事を選択し，10 トピックからなる計二十記事のデータセットを作成した．このとき，10 トピックのうち，5 トピックに含まれる二つの記事は，その記事の報道元が属する国が同一であるように選択した．また，他の 5 トピックの記事は，その記事の報道元が属する国が異なるように選択した．

ユーザ評価は四人の大学生を対象に行った．各々の評価者は十個のそれぞれのトピックに対し，そのトピックに属する二つの記事と，その二つの記事に対応する二つの分析結果を参考にしながら記事を読み比べる．読み比べた後，アンケートに回答する．

アンケートは十個のトピックのそれぞれに対応する以下の二つの質問から構成される．

質問 1　マイニング結果は記事の偏りの理解に役立つか．

質問 2　マイニング結果は記事の偏りを表しているか．

各質問に対し 1 から 5 の評点で評価を行う．5 が最も良い評価であり，1 が最も悪い評価とする．

各質問に対する評価結果を図 5，図 6 に示す．各図はそれぞれの質問に対して，同一国の記事から構成されるペアに対する評点の割合，異なる国の記事から構成されるペアへの評点の割合，両方を合わせた評点の割合の三つを表す．また，図 6 はそれぞれの質問に対する評点の平均と，評点が 5 または 4 であったペアの割合である．

5.4 考　　察

表 5 で記した様に，記事から抽出された利害関係の適合率，再現率は 65% から 70% であった．このことから，記事に記述されている利害関係を高い精度で抽出できていることがわかる．また，ステークホルダーに対する記述極性の分析の精度は 67.3% であり，高い精度で分析可能であると言える．しかし，記事の報道元や解説を求められた専門家など，利害関係を持たない実体もステークホルダーとして抽出されていることがあった．また，今回の実験では，関係語及び感情語の文脈に即した意味を考慮していない．よって，さらなるマイニング手法の改善が必要であると考えられる．

マイニング結果が記事の偏りを表していることは，ユーザ評価において良い評価を得ていることからわかる．表 6 に示す通り，質問 1 において偏りの理解に役立つという評価の割合は 58.5%，質問 2 において分析結果が偏りを表しているという評価の割合が 70.0% であった．よって，ステークホルダーマイニングはニュース記事の偏向分析に有用であることがわかる．

偏向分析の結果における同国のペアと異国のペアとの比較では明らかに違いが見られた (表 6)．いずれの質問においても，報道元が異なる国に属す記事の組の方が偏りが表れており，偏りの理解ができることがわかる．特に，質問 1 と質問 2 は共に 75.0% の評点が 5 及び 4 であり，偏りが大きいことが表れている．また，同国の記事であっても質問 2 においても 65.0% の評点が 5 または 4 であったため，同国の記事であっても偏りを分析できている．

以上の考察から，ステークホルダーマイニングはニュース記事の偏向分析に有用であると結論できる．

6. 結　　論

本稿ではステークホルダーマイニングと，その結果を利用したニュース報道の偏向分析支援手法を提案した．

ステークホルダーマイニングでは，我々は記事を分析し，記事に記述されているステークホルダーとステークホルダー間の利害関係を抽出した．また，利害関係の抽出のために，関係語辞書と関係構文構造を作成した．さらに，抽出されたステークホルダーに対する記述極性の分析を行った．そして，ステークホルダーマイニングによるニュース記事の分析結果を用いることにより，ニュース記事に対する偏向の客観的基準を提示し，偏向分析を支援する手法を提案した．

評価実験では，マイニング結果による偏向分析の有用性が示せた．マイニング結果を用いることにより，事象に対する記事の視点の定量化，記事の偏りを補完する記事の発見，記事の偏りの大きさの数値化などといった応用が考えられる．

今後の課題としてステークホルダーマイニングの手法の改善が挙げられる．現在の手法では，記事から抽出した固有表現について，エンティティの同一判定を行っていない．例えば米国は “America” 以外に “U.S.”，“USA” など複数の表記があが，これは本来同一の実態を表す．よって同一判定を行うことでさらなる精度の改善が見込めると考えられる．また文脈を考慮した辞書の利用が必要である．

謝辞　本研究の一部は，文部科学省科学技術研究費（20700084 と 20300042）の助成を受けたものである．

文　　献

[1] J. Liu and L. Birnbaum: “Localsavvy: aggregating local points of view about news issues”, Proceedings of the first international workshop on Location and the web, pp. 33–40 (2008).

[2] S. Park, S. Kang, S. Chung and J. Song: “Newscube: delivering multiple aspects of news to mitigate media bias”, Proceedings of the 27th international conference on Human factors in computing systems, pp. 443–452 (2009).

[3] A. Nadamoto and K. Tanaka: “A comparative web browser (cwb) for browsing and comparing web pages”, Proceedings of the 12th international conference on World Wide Web, pp. 727–735 (2003).

[4] G. A. Miller: “Wordnet: a lexical database for english”, Commununications of ACM, **38**, 11, pp. 39–41 (1995).

[5] A. Esuli and F. Sebastiani: “Sentiwordnet: A publicly available lexical resource for opinion mining”, Proceedings of the 5th Conference on Language Resources and Evaluation, pp. 417–422 (2006).

[6] Q. Ma and M. Yoshikawa: “Topic and viewpoint extraction for diversity and bias analysis of news contents”, Advances in Data and Web Management (Eds. by Q. Li, L. Feng, J. Pei, S. X. Wang, X. Zhou and Q.-M. Zhu), Vol. 5446, chapter 15, pp. 150–161 (2009).

[7] S. Ishida, Q. Ma and M. Yoshikawa: “Analysis of news agencies’ descriptive features of people and organizations”, Database and Expert Systems Applications (Eds. by S. S. Bhowmick, J. Küng and R. Wagner), Vol. 5690, chapter 62, pp. 745–752 (2009).

[8] G. Grefenstette, G. Grefenstette, Y. Qu, J. G. Shanahan and D. A. Evans: “Coupling niche browsers and affect analysis for an opinion mining application”, Proceedings of the 7th international conference on Adaptivity, Personalization and Fusion of Heterogeneous Information, pp. 186–194 (2004).

[9] Google News: `http://news.google.com/`.

[10] M. Marneffe, B. Maccartney and C. Manning: “Generating typed dependency parses from phrase structure parses”, Proceedings of the 5th Conference on Language Resources and Evaluation, pp. 449–454 (2006).

[11] M. C. De Marneffe and C. D. Manning: “The Stanford typed dependencies representation”, Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation, pp. 1–8 (2008).