

# メディア情報の Linked Data 化と活用事例の提案

川村 隆浩<sup>†</sup>      越川 兼地<sup>†</sup>      中川 博之<sup>†</sup>      清 雄一<sup>†</sup>  
田原 康之<sup>†</sup>      大須賀昭彦<sup>†</sup>

## Proposal of Social-Mass Media Triplification and Its Use Case

Takahiro KAWAMURA<sup>†</sup>, Kenji KOSHIKAWA<sup>†</sup>, Hiroyuki NAKAGAWA<sup>†</sup>, Yuichi SEI<sup>†</sup>,  
Yasuyuki TAHARA<sup>†</sup>, and Akihiko OHSUGA<sup>†</sup>

あらまし 昨今、インターネットの普及などから様々な情報源（ソーシャルメディア・マスメディア）に容易にアクセスし、多様な意見・考え方に触れることが可能になった。同時に、ソーシャルメディアにおけるデマの拡散や、マスメディアにおける偏向報道・情報操作の疑いなど、ユーザ自身が情報の信頼性について自ら判断することが求められてきている。そこで我々は、一般ユーザがメディア情報を多角的な観点から比較することを支援するため、ユーザに代わってソーシャル、マス両メディアから特定の話題に関する情報を抽出、見える化し、特定の観点に基づく比較ポイントを提示するエージェントシステムを目指している。本論文では、Conditional Random Fields と事象抽出のためのヒューリスティクスを用いて、Twitter 上のツイート、マスメディアのニュース記事等から 13 の属性情報をもつ事象情報を抽出し、それらを Linked Data 化する手法を提案し、精度評価を行った。また、事例を通して多様性、希少性、偏在性、因果関係の四つの観点に沿って比較ポイントを抽出することで有用性を確認した。

キーワード エージェント, Linked Data, メディア比較

### 1. ま え が き

近年、インターネットを通して様々な情報源（ソーシャルメディア・マスメディア）に容易にアクセスできるようになった。特にソーシャルメディアの普及は凄まじく、多様な意見・考え方にインターネットさえあれば容易に触れることが可能になった。しかし、Twitter 上でのデマ拡散の問題や、従来のマスメディアにおいても偏向報道や情報操作に関する疑いの声が一般ユーザにまで届いてきている。そのため、情報収集の本来のあり方である多角的な観点からの入手と、自らの判断で理解することがますます重要になってきている。

そこで我々は、ユーザに代わって両メディアから情報を抽出（ただし、文字情報に限定）、見える化し、比較・探索を可能とするエージェントサービスの実現を目指している。本論文では、ニュース記事、ツイートを対象に、特定の話題に関する事象情報（5W1H 等）

を条件付き確率場（Conditional Random Fields, 以下 CRF）とヒューリスティクスを用いて抽出し、それらを Linked Data 化 [1] することで事象ネットワークとして可視化する手法を提案する。Linked Data とは、インターネットを介したデータ共有方法として近年、注目を集めている技術である [2], [3]。RDF (Resource Description Framework) 記述内で、URI (Uniform Resource Identifier) を用いて Resource を一意に表現し、それらをセマンティクスが定義された Property を用いてリンクするものであり、様々な情報源（データセット）のデータを容易に結び付けることができる。また、事象ネットワークの活用事例としてまず輸送機オスプレイ導入に関する Linked Data の比較を行い、メディア比較における観点を定義し、それらの観点に沿って 2012 年衆議院議員総選挙・東京都知事選挙に関する Linked Data から注目すべきポイントを SPARQL 検索により自動的に抽出、ユーザに提示することで、有効性を評価する。

本論文の構成を以下に示す。まず 2. では、メディア情報から Linked Data への変換方法と変換精度について述べる。次に 3. では、オスプレイに関するメデ

<sup>†</sup> 電気通信大学大学院情報システム学研究所, 調布市  
Graduate School of Information Systems, University of  
Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi,  
182-8585 Japan

ア比較を通じて比較観点を洗い出し、総選挙に関するメディア比較へ適用した事例を示す。最後に、4. で関連研究を示し、5. でまとめと今後の課題について述べる。

## 2. メディア情報の Linked Data 化手法

本論において“事象”とは、実世界の出来事を表す情報（例：「日経平均株価が3日連続で上昇している.」, 「安倍氏と東京駅で握手をした.」など）を意味している。これら事象に関する情報は各メディア（テレビ・新聞・ニュースサイト・ブログ・Twitter など）を通じて我々のもとに伝えられ、日々の行動を決める際の重要な情報となっている。そこで本研究では、これら事象情報を計算機が扱いやすい、整理された形式にするため、自然文で表現された事象情報（ニュース記事・ツイートなど）をセマンティクスを考慮した構造化された形で再表現する (Linked Data 化)。具体的には、自然文から事象属性（次節参照）を抽出し、意味ネットワーク（各エッジは抽出された事象属性、各ノードは事象属性に対応するキーワード）を形成する。これによりもとの自然文の意味を構造化された情報として表現することができると考えている。なお、提案手法は Nguyen らの先行研究 [4] をベースに以下の拡張を行ったものである。

### 2.1 事象属性の追加

事象情報を構造化された情報として表現するために、Nguyen らが定義した行動属性を拡張し、新たに13種の事象属性を定義した。Nguyen らは、ユーザの行動情報を表現するために行動属性と行動間の遷移を {行動主: Who, 動作: Action, 対象: What, 場所: Where, 時刻及び場面: When, 行動間の遷移ラベル次: Next, 後: After, 行動間の因果関係: Because of} として定義した。しかし、ユーザ行動の抽出を目的としていたため、人間の行動を含まない文は対象としていなかった。それに対し、本論では人間の行動を表さない内容（例：「京王線 新宿～調布間・運転見合わせ中。調布駅混んでる。」など）も対象とするため、ニュース記事等を分析し、事象を表現する上で必要となる属性を Nguyen らの行動属性を拡張する形で表 1 のように定義した。ただし、今回、対象を見聞した内容にも拡大したことで、事象の発生場所 (Location) は必ずしもユーザの位置情報 (Where) と同義ではない。事象がユーザ行動であれば、事象の発生場所とユーザの位置情報は一致するため、後方互換であるが、ユーザ

行動でない事象に発生場所に関する情報があり、かつ、ユーザ（発信者）の位置情報が付いていた場合は、事象の発生場所（事象内の位置情報）を優先する。

追加した属性の使用例として、人間の行動を表さない事象を含む以下の例文をネットワークで表現した例を図 1 に示す（Status 属性については次節の例を参照して欲しい）。

例文：「静岡日報によれば、台風になれば大井電鉄のアルプスライナーは運行を休止することを 16 日、記者団に発表した」

### 2.2 CRF を用いた事象属性の推定

自然文から事象情報を抽出し、Linked Data として出力する本手法の流れを図 2 に示す。なお、対象言語は日本語である。ここでは、まず前処理として Twitter 特有の文字列（リプライ/ユーザ名/ハッシュタグ/URL/RT/QT）の置換や形態素解析辞書への登録を行った後、形態素解析、構文解析を掛け、その

表 1 事象属性一覧  
Table 1 List of event attribute.

事象属性	意味	先行研究との対応
Subject	主題	Who
Activity	動作	Action
Object	動作の目的語	What
Time	事象が起こる時刻及び場面	When
Location	事象が起こる場所	Where
Cause	事象が起こる原因	BecauseOf
Next	ある事象の次の事象	Next
Target	動作の対象主 (動作を誰に向けて行ったか?)	追加
Status	主題の状態	追加
Quoted source	情報の発信元	追加
Regard	事象を捉える立場/観点	追加
Modifier	修飾句	追加
Case	事象が起こる条件	追加

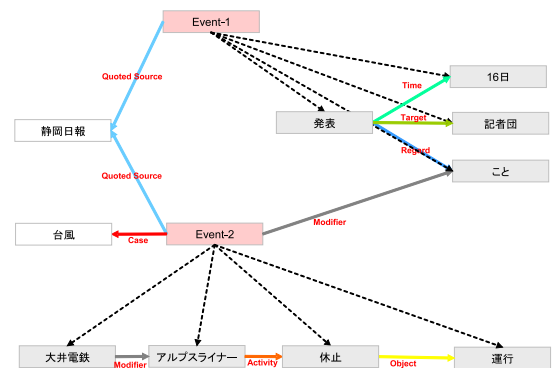


図 1 事象ネットワークの例  
Fig.1 Example network.

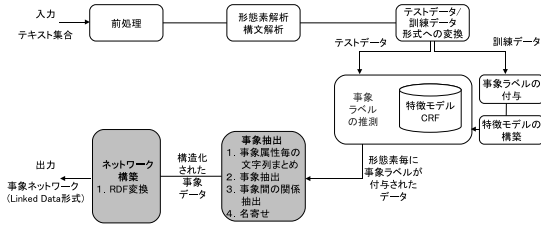


図 2 事象情報抽出の流れ  
Fig. 2 Flow of event extraction.

結果を CRF に通して事象属性を推測している。なお、形態素解析器には MeCab<sup>(注1)</sup>、構文解析器には CaboCha<sup>(注2)</sup>を用いた。

本節では、先行研究 [4] である CRF を用いた事象属性の推定手法の概要を示す。CRF とは、John D. Lafferty ら [5] が提案した系列ラベリング問題に適用するための識別モデルである。文字種や周辺の単語情報などといった素性を素性関数という形で柔軟に投入でき、Hidden Markov Models や Maximum Entropy Markov Models の問題点 (label bias, length bias) を自然にかつ有効に解決できる。これまでに品詞付与 [5]、テキストチャンキング [6]、固有表現抽出 [7]、形態素解析 [8] などといった系列ラベリング問題に適用され、いずれにおいても高い精度を示している。

ここではまず CRF を用いて特徴モデルを生成するために対象文に形態素解析、構文解析を掛けた後、以下のデータ形式に変換する。

- (1) データ元参照情報
- (2) 鍵括弧の対応付け情報
- (3) 文節情報 (文節 ID/係り受け先文節 ID)
- (4) 表層形
- (5) 品詞 ID

**データ元参照情報**とは、原文が保存されているデータベースへの参照情報を示しており、事象ネットワーク構築後に事象と原文との紐付けを行う際に用いる情報である。

**鍵括弧の対応付け情報**とは、事象属性と文中の対応文字列を同定する過程において、ニュース記事で頻出する鍵括弧 (「」, 『』, 【】 など) を含んだ文に対して適切に処理するため、鍵括弧の内部の文は別に処理している。事象ネットワーク構築後に鍵括弧が存在した文から得た事象と鍵括弧内部の文から得た事象を紐付け

るために用いる情報である。

文節情報、表層形、及び品詞 ID は、形態素解析、構文解析した結果から得ることができる。

次に、形態素ごとに前節の事象属性を人手で付与 (ラベリング) し、訓練データを作成する。事象属性ラベルは、文字「B」、「I」と各事象属性の組合せと文字「O」、「M」で表現し、それぞれの頭文字は「B (Begin, 表現の始まり)」、「I (Inside, 表現の途中)」、「O (Outside, それ以外)」、「M (Modifier 修飾句)」を意味している。そして、訓練データに対してどのような素性を与えるかを定義するためのテンプレートファイルを作成する。本手法が利用する素性は表層形、品詞 ID、事象属性の三つである。テンプレートファイルはこれらの素性を使った特徴を定義するものであり、ここでは長文に対応させるため、サイズ 7 (対象語から前後 3 語まで) のウィンドウを採用した。

このテンプレートファイルと訓練データを用いて、特徴モデルを構築し、構築した特徴モデルを使って、テストデータの事象属性の推測を行う。表 2 に CRF の出力例を示す。

### 2.3 事象抽出のためのヒューリスティックルールの導入

次に、CRF の出力 (形態素ごとに事象属性が付与された情報) から文節情報や品詞情報を手掛かりとしたヒューリスティックルールを用いて“事象”を抽出する。ここで事象とは、一つ以上の事象属性をエッジとし、事象属性の Activity (行動)、Status (状態) を中心とした意味のまとまりをもつ部分グラフである。

CRF の出力形式から事象を抽出するヒューリスティックルールはまず、各事象属性に対応する文字列の連結 (2.3.1)、次に、意味的関連がある事象属性で連結することで事象を抽出し (2.3.2)、最後に、事象間の関係を抽出 (2.3.3) することで、CRF の出力形式 (例: 表 2) から構造化された事象情報 (例: 表 4) を得る。以下にこの三つの工程について説明する。

#### 2.3.1 各事象属性に対応する文字列の連結

CRF の出力結果から事象を抽出するにあたり、まず CRF の出力形式 (入力例: 表 2) から各事象属性に対応する文字列を連結し、修飾句の修飾先の事象属性を同定する必要がある (出力例: 表 3)。

そこで、事象属性ラベルに付与されている先頭の頭文字「B」、「I」、「O」、「M」に着目し、以下のヒューリスティックルールを用いて各事象属性に対応する文字列を連結する。

(注1) : <https://code.google.com/p/mecab/>

(注2) : <http://code.google.com/p/cabocha/>

表 2 CRF 出力例  
Table 2 Example of CRF output.

データ元 参照情報	鍵括弧の 対応情報	文節ID	係り受け先 文節ID	表層形	品詞ID	事象属性 ラベル
329-10	0	0	1	半年	67	M
329-10	0	0	1	間	51	M
329-10	0	0	1	の	24	M
329-10	0	1	2	運航	36	B-Object
329-10	0	1	2	休止	36	I-Object
329-10	0	1	2	を	13	O
329-10	0	2	4	余儀なく	10	B-Activity
329-10	0	2	4	さ	31	I-Activity
329-10	0	2	4	れ	32	I-Activity
329-10	0	2	4	た	25	I-Activity
329-10	0	3	4	大きな	68	O
329-10	0	4	5	要因	38	B-Subject
329-10	0	4	5	が	13	O
329-10	0	5	-1	高速料金	38	B-Status
329-10	0	5	-1	だ	25	O
329-10	0	5	-1	。	7	O

表 3 事象属性ごとの文字列  
Table 3 Event attribute strings.

事象属性	対応文字列
M-Object	半年間の
Object	運行休止
Activity	余儀なくされた
Subject	要因
Status	高速料金

(1) 抽出する属性、及び対応する文字列は BIO タグが B で始まる単独の形態素、若しくは複数の連続した形態素で構成される。

(2) 修飾句は単独若しくは連続する形態素で構成される。

(3) 修飾句は自身の修飾句の次の形態素から始まる属性を修飾する。

このルールを例：表 2 に適用すると、表 3 のように各属性に対応する文字列を得ることができる。表 3 の“M-Object”という事象属性は直後の事象属性 Object を修飾する (Modifier) ことを意味している。

### 2.3.2 事象ごとの事象属性の連結

前工程では各事象属性に対応する文字列を連結したが、一文に複数の事象が存在することがあるため、必ずしも前工程で得られた結果を一事象としてみなすことはできない。そこで、文意を踏まえた上で事象ごとに意味の関連がある事象属性を連結し、各事象を構成するため、文節間の関連情報である係り受け解析の結果と事象属性の Activity (行動)、Status (状態) に着目し、以下のような事象ごとに事象属性を連結するルールを導入した。

(1) 事象属性 Activity/Status が出現する文節を文意の区切りとみなし、事象属性 Activity/Status が出現する文節 ID ( $e\_id$  とする) のリストを作る。なお、末尾の文節 ID がリストに含まれていない場合は

加える。以降、本リストを  $e\_list$  とする。

例： $e\_list \Rightarrow [2,5]$  (対応文字列：“余儀なくされた”，“高速利用金だ。”)

(2)  $List$  を文節 ID の昇順に並べ換える。要素の  $e\_id$  に着目し、その  $e\_id$  を係り受けしている文節 ID を連結する。なお、既に他の  $e\_id$  として連結した文節 ID は含めないものとする。

例： $e\_id: 2 \Rightarrow [0,1,2]$  ( $event1$  とする),  $e\_id: 5 \Rightarrow [3,4,5]$  ( $event2$  とする)

(3)  $e\_id$  ごとに束ねられた文節 ID に含まれる事象属性の組合せを一事象とみなし、各事象を抽出する。  
例： $event1 [0,1,2]$  中に含まれる事象属性  $\Rightarrow$  [M-Object: “半年間の”, Object: “運行休止”, Activity: “余儀なくされた”]  $event2 [3,4,5]$  中に含まれる事象属性  $\Rightarrow$  [Subject: “要因”, Status: “高速料金”]

### 2.3.3 事象間の関係を抽出

前工程では意味的関連のある事象属性を連結することで、事象を抽出した。この段階で事象情報は構造化されているが、事象情報をよりリッチな情報にすべく事象間の関係抽出を試みる。

一般に実世界に起こる事象は他の大多数の事象と相互作用しあいながら生じるものであることが知られている。Radinsky らは New York Times<sup>(注3)</sup> の 150 年間の過去のニュース記事を対象に事象情報を抽出し、因果関係に着目した予測木を作ることで、ある一定の条件において人間よりも高い精度で未来に起こる事象を予測できることを示した [9]。

Radinsky らの研究で事象間の関係の有効性を示したように、メディア比較・分析のタスクにおいても、事象間の関係情報が重要な情報になると考え、事象間の関係を抽出する。そこで、比較的多く観測できた事象間の関係を表す修飾句、目的語句、要因句、条件句、並列句、主語句を同定するヒューリスティックルールを導出した。

前項で定義した事象属性 Activity/Status が出現する文節 ID である  $e\_id$ 、 $e\_id$  の係り受け先文節 ID を  $link\_id$ 、 $e\_id$  の次の文節 ID を  $next\_id$  とする (次の文節が存在する場合、 $next\_id = e\_id + 1$  が成り立つ)。  
・修飾句になる条件

$link\_id = n\_id$  であり、かつ、 $e\_id$  の末尾の形態素と対応する事象属性が Activity であり、かつ、 $e\_id$  の品詞が“名詞”または“接頭詞”である場合、 $e\_id$  が属す

(注3) : <http://www.nytimes.com/>



る事象は *next\_id* の先頭の語の修飾句である。

例文：父が購入したバットを試合で使った。

*event*: (修飾句 **M-Object**: (Subject: 父, Activity: 購入する),  
Object: バット, Location: 試合, Activity: 使う)

・目的語句となる条件

*link\_id* = *n\_id* であり、かつ、*e\_id* の末尾の形態素と対応する事象属性が Activity であり、かつ、*e\_id* の品詞が“動詞”である場合、*e\_id* が所属する事象は *next\_id* の先頭の動詞の目的語句である。

例文：天気が回復するかわからない。

*event*: (目的語句 **Object**: (Object: 天気, Activity: 回復する),  
Activity: わからない)

・要因句となる条件

*link\_id* に Activity 属性が存在し、かつ、*e\_id* の末尾の文字列が因果関係の要因を表す手がかり表現（後述）に該当する場合、*e\_id* が属する事象は *link\_id* が属する事象の因果関係句である。

例文：公務を怠ったとして、減給処分になった。

*event\_A*: (Object: 公務, Activity: 怠る)

*event\_B*: (Object: 減給処分, Activity: なる)

*event\_A* は *event\_B* の要因句である。(*event\_B* - *cause* → *event\_A*)

・条件句となる条件

*link\_id* に Activity 属性が存在し、かつ、*e\_id* の末尾の文字列が条件句を表す手がかり表現（後述）に該当する場合、*e\_id* が属する事象は *link\_id* が属する事象の条件句である。

例文：目的税にした社会福祉のためにつかうのであれば国民も納得するだろう。

*event\_A*: (M-Object: (Object: 目的税, Activity: する), Object: 社会福祉, Activity: つかう)

*event\_B*: (Subject: 国民, Activity: 納得する)

*event\_A* は *event\_B* の条件句である。(*event\_B* - *case* → *event\_A*)

・並列句となる条件

*link\_id* に Activity 属性が存在し、かつ、*e\_id* が属する事象が因果関係句である条件にも条件句である条件にも該当しない場合、*e\_id* が属する事象と *link\_id* が属する事象の関係は並列である。

例文：市民 6 人が 3 日、市長給与の返還などを求め、市に監査請求した。

*event\_A*: (Subject: 市民 6 人, Time: 3 日, M-Object: 市長給与, Object: 返還, Activity: 求める)

*event\_B*: (Subject: 市民 6 人, Time: 3 日, Target: 市, Activ-

ity: 監査請求する)

*event\_A* と *event\_B* は並列句であり、*event\_A* の後に *event\_B* が起こる。(*event\_A* - *next* → *event\_B*)

・主語句となる条件

*link\_id* に Status 属性が存在し、かつ、*e\_id* の末尾の文字列が係り受け先の主語句となる手がかり表現（後述）に該当する場合、*e\_id* が属する事象は *link\_id* が属する事象の主語句である。

例文：景気対策をやるのが前提だ。

*event*: (主語句 **Subject**: (Object: 景気対策, Activity: やる), Status: 前提)

なお、因果関係句を表す手がかり表現は { なり／ても／には／に伴う／おらず／生じて／上で／ため／ために／として／により／によって／が受けて／の効果／が／を背景に／の影響も／の影響が } の 17 種、条件句を表す手がかり表現は { あれば／のに／場合／すれば／ければ／けれども } の 6 種、係り受け先の主語句となる手がかり表現は { のが／のは／のも } の 3 種とした。上記のルールを適用することで、先の結果から表 4 の形で事象情報を得ることができる。

最後に、事象属性に対応する文字列を主語・目的語に、事象属性を述語に対応させることで Linked Data におけるトリプル {主語, 述語, 目的語} を構成する。トリプルへの変換ルールを表 5 に示す。また、上記例に適用し、事象ネットワークを構築した例を図 3 に示す。なお、動詞等は原型に戻し、日本語シソーラス辞書 WordNet<sup>(注4)</sup>を用いて可能な限り同義のノードを名寄せする。

## 2.4 事象抽出精度の評価

前節で提案したヒューリスティックルールの有効性を検証するため、事象抽出の精度を評価した。まず訓練データ (Twitter データ, 記事データ) に対し、人手で事象ラベルを付与し、内 9 割を訓練データ, 1 割をテストデータとした 10 交差検定を行い、事象属性ラ

表 4 抽出された事象情報  
Table 4 Event information extracted.

Event ID	事象属性	対応文字列
329-10-1	M-Object	半年間
	Object	運行休止
	Activity	余儀なくされる
329-10-2	M-Subject	329-10-1
	Subject	要因
	Status	高速料金

(注4) : <http://nlpwww.nict.go.jp/wn-ja/>

表 5 トリプル変換ルール  
Table 5 Triplification rules.

関係名	属性	対象属性	トリプル例	組合せ数
基本属性 (主題)	Subject	Activity, Status	$(Subject_{str}, Activity, Activity_{str}),$ $(Subject_{str}, Status, Status_{str})$	2
基本属性 (動作)	Activity	Object, Location, Time, Target, Regard	$(Activity_{str}, Object, Object_{str}),$ $(Activity_{str}, Time, Time_{str}),...$	5
基本属性 (状態)	Status	Location, Time, Regard	$(Status_{str}, Location, Location_{str}),$ $(Status_{str}, Time, Time_{str}),...$	3
事象間のつながり (因果関係, 条件, 並列, 引用)	Event_id	Cause, Case, Next, Quoted_source	$(Event\_id_{str}, Cause, Cause_{str}),$ $(Event\_id_{str}, Quoted\_source,$ $Quoted\_source_{str}),...$	4
データ元参照	Event_id	Subject, Time, Status, Object, Location, Time, Target, Regard, (Cause, Case) <sup>(注5)</sup>	$(Activity_{str}, Activity, Event\_id_{str}),$ $(Time_{str}, Time, Event\_id_{str}),...$	10

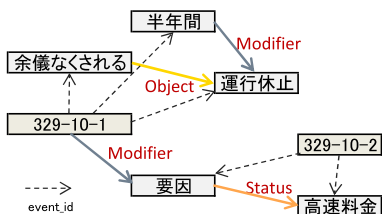


図 3 事象ネットワークの例  
Fig. 3 Example network.

ベルの推測精度を算出した。訓練データの概要を表 6 に示す。個々の事象の正解基準は、正解データとして用意した事象属性（エッジ）とキーワード（ノード）の組みが全て一致するかどうかである。表 7 に結果を示す。なお、表中、“先行研究”は 13 種の属性を扱うものの、ヒューリスティックルールは導入していない手法を指す。また、“事象間関係”の抽出は先行研究では行っていない。

この結果から、先行研究に比べ、より高い精度でラベルを推定できることを確認した。これは、CRF を用いていわば機械的にラベルを推定していた先行研究に対し、今回、“事象”という概念を導入し、事象の単位でまとめ、事象間の関係を明確化するヒューリスティックルールを導入したことで、文の基本構造をより正確に捉える効果があったものと思われる。しかしながら、推測精度は属性ごとにまだばらつきがあり、基本属性となる Activity や Object, Modifier などは F 値約 80% で推定できているのに対し、Regard, Cause, Quoted source などは 30% を下回っていた。これらはいずれも今回、導入した属性であり、更なる洗練化が求められる。

また、属性によっては Twitter と記事データで精度に差があり、Location, Time, Quoted source は記

表 6 実験データの概要  
Table 6 Experimental results.

訓練データ	文の数	ラベル数	modifier	subject	activity	object
ニュース記事	119	2,232	320	162	423	232
Twitter	175	2,529	254	240	489	253
ニュース記事 + Twitter	294	4,761	574	402	912	485

訓練データ	target	status	location	time	regard	cause	quoted source
ニュース記事	12	39	82	80	37	18	14
Twitter	12	159	47	74	34	5	10
ニュース記事 + Twitter	24	198	129	154	71	23	24

表 7 CRF 出力からの事象・事象間関係の抽出精度  
Table 7 Event extraction accuracy from CRF output.

	事象抽出		事象間関係
	提案手法	先行研究	抽出
適合率 (%)	74.7	53.9	97.4
再現率 (%)	74.7	44.3	79.2
F 値	74.7	48.6	87.4

事データでは高精度だが Twitter では低く、反対に、Cause は Twitter では高精度だが、記事データでは低い結果となった。これは属性ごとに表現の多様性に差があり、ランダムに収集した学習データで十分にカバーできていなかったことが原因と思われる。

更に、事象間の関係を推測することができるようになったことは、次章で示すメディア比較に多いに有用と思われる。

### 3. メディア情報の比較事例

#### 3.1 メディア比較における観点の検討

本章では、まず予備実験として輸送機オスプレイの導入問題<sup>(注6)</sup>に関して、ソーシャル、マス両メディア

(注6)：存日米軍基地に現行の旧型輸送機に代わり新型輸送機オスプレイを搬入しようとする/した動きに対して、オスプレイの安全性などの観点で導入することに対して賛否がわかれている問題を指す。

から集めた情報を対象に、Linked Data 化を行い、両者の比較を行う [10]。

ソーシャルメディアの対象としては、マイクロブログサービス最大の Twitter を選定し、一定期間 Streaming API にて取得したツイートからユーザーリストを作成し、そのユーザーリストから無作為にユーザーを選択、順次 REST API にて過去ツイートを遡って取得した。その上で、事前に作成したオスプレイに関するキーワードリストを用いてツイートの絞り込みを行った。マスメディアとして、全国紙を発行する新聞社 5 社、及びキー局 5 社に日本放送協会 (NHK) を加えたテレビ局 6 社が運営するニュースサイトに投稿された記事を対象とし、各サイトの最新記事一覧または検索機能から記事ページを辿り、各サイトごとに作成したスクレイピングプログラムを用いて記事本文を取得した。データの収集期間はいずれも 2012/04/01～2012/08/16 とし、約 3000 ツイート (絞り込み後)、535 記事を収集した。

その上で、rdf-gravity<sup>(注7)</sup>や Gephi<sup>(注8)</sup>を用いて、ネットワークを可視化した。ノードの色、及びエッジの色の意味を表 8 に示す。ネットワークの一部を図 4 に示す。この図から「配備」ノードから「オスプレイ」ノードへ Object 属性で繋がっており、また「配備」ノードから「米軍普天間飛行場」・「沖縄」ノードにも Object 有向エッジが伸びていることから、オスプレイが沖縄若しくは普天間飛行場に配備される (若しくはされたことが) ネットワークから読み取れる。また、「配備」に「反対」していること声があることや、紫色のノードはソーシャルメディアでのみ出現する語を意味しているため、沖縄ローカル紙である琉球新報からの情報はソーシャルメディアでしか得られなかったことが読み取れる。以下では、ソーシャル、マス両メディアによるネットワークを人手により比較し、注目すべき観点の洗い出しを行った。

### 3.1.1 事象の多様性

表 9 における各ネットワークを構成するノード数<sup>(注9)</sup>、及び両メディアのネットワーク図 5、図 6 か

表 8 ノード、エッジの色  
Table 8 Colors of nodes and edges.

ノードにおける マスメディア の構成比率	色	エッジの種類		色
0%	紫	Activity		赤
37.5 % 未満	青	Object		黄
62.5 % 未満	緑	Time		緑
100 % 未満	黄	Location		青
100%	黄	Cause		赤
		Target		青
		Status		黄
		Quoted source		黄
		Event_id		黄

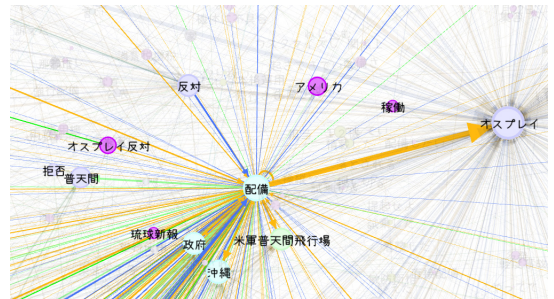


図 4 オスプレイに関する事象ネットワーク (配備ノードに着目)  
Fig.4 Network for Osprey problem.

ら空間に対するノード数の密度の違いが読み取れる。更に、図 7 は Twitter と新聞紙メディアの両者の事象ネットワークを可視化した様子の外観である。紫色のノードはソーシャルメディアでのみ出現する語を、黄緑色は新聞紙メディアでのみ出現する語を表している。図では明らかに紫色のノードが多く、マスメディアではイベントや事件に対する考え方・意見がある程度、統一化されているのに対し、ソーシャルメディアでは多くの一般ユーザーの多種多様な意見が含まれていることが確認できる。一方、新聞紙メディアでのみ登場する語としては、新聞は事実関係を正確に伝えようとすることから、正確な時間や場所、アメリカ側の人名や正式な役職 (～報道官、～政務官など) などが散見された。また、一部の遠回しな表現 (忙殺されて、言及されていないなど) に関しては、ソーシャルメディア側の平易な同義語との名寄せに失敗しているものが見られた。

図 5、図 6、図 7 の拡大版を著者のサイト<sup>(注10)</sup>に示す。

### 3.1.2 希少事象へのアプローチ

一般的に頻度は意思決定の重要な判断材料になり得

(注7) : <http://semweb.salzburgresearch.at/apps/rdf-gravity/index.html>

(注8) : <http://oss.infoscience.co.jp/gephi/gephi.org/index.html>

(注9) : ソーシャルメディアでは対象 (目的語など) が陽に記載されていない書込みが散見されるため、Target の抽出数が他のメディアに比べ少ない。また、Status は三つの基本属性 (表 5) の一つであり、主題でも動作でもないものがここに分類されるため (be 動詞に続く補語など)、ソーシャルメディアにおける話題の多様性が最も表れている。

(注10) : <http://www.ohsuga.is.uec.ac.jp/~kawamura/fig5-7.pdf>

表 9 オスプレイに関する事象ネットワークに関するデータ  
Table 9 Network information for Osprey problem.

メディア	ノード数	得られた 事象数	ラベル計	Subject	Activity	Object	Target	Status	Time	Location	Cause	Quoted source
Twitter	4,218	6,846	15,703	2582	6129	4636	16	1130	520	548	13	129
MSN産経ニュース	2,134	3,790	9,669	1258	3733	3277	174	91	554	565	16	1
朝日新聞デジタル	1,339	1,920	5,003	706	1894	1650	102	48	255	332	14	2
日テレ NEWS24	294	409	1,198	145	408	340	12	1	139	149	3	1
FNN	917	1,712	4,148	564	1666	1387	71	10	149	292	9	0

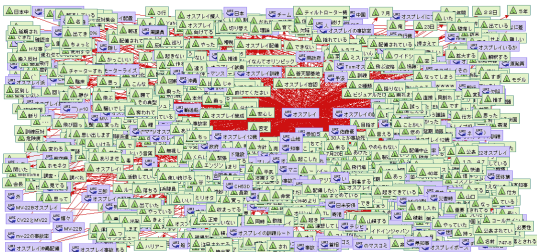


図 5 Twitter から構築した事象ネットワーク  
Fig.5 Network from Twitter.

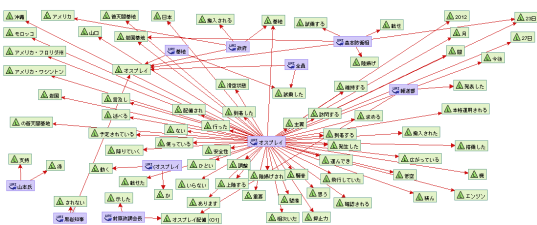


図 6 FNN から構築した事象ネットワーク  
Fig.6 Network from FNN.

るが、場合によっては少数の意見の中に重要な情報が含まれており、頻度分析ではこれらマイナーな意見はノイズとして判断され、取りこぼされてしまうことがある。したがって、本手法ではマイナーな意見も最終的なネットワークに現れるように配慮している。図 8 は Twitter から構築したネットワークの一部である。この図には、マスメディアにはないオスプレイ配置に対する賛成意見が現れている。

更に、Twitter から構築した事象ネットワークの一部 (図 9, 10) に注目すると、マスメディアにはないノード「MV-22」, 「CV-22」がソーシャルメディアに顕著に現れていることが分かる。これらはオスプレイの型番であり、MV-22 (正確には MV-22B) は主に輸送を目的とした機種 (事故率 1.93), CV-22 (正確には CV-22B) は特殊作戦を目的とした機種 (事故率 13.47) を表している。じつは、日本に配備されるのは輸送型 MV-22 であり、約 40 年前から配備されている現行 CH-46 ヘリコプターの事故率 2.43 より低い。

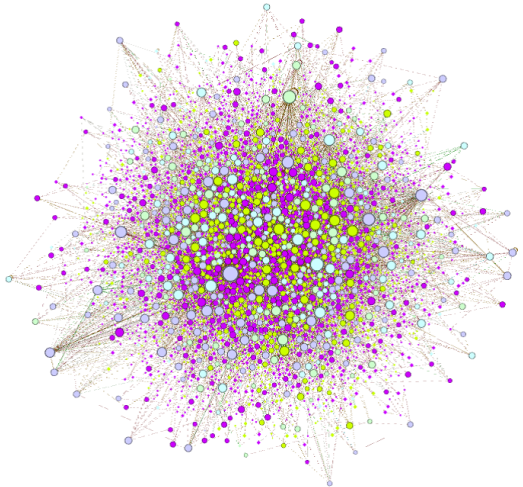


図 7 オスプレイに関する事象ネットワークの外観  
Fig.7 Network overview for Osprey problem.

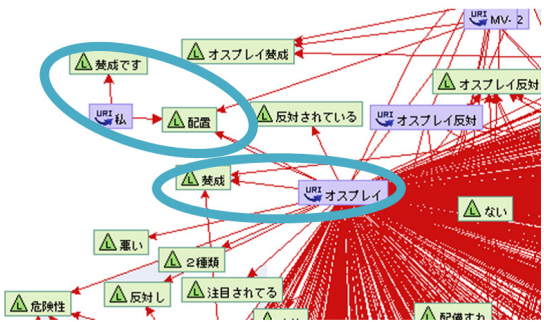


図 8 オスプレイに関するサブネットワーク (賛成意見)  
Fig.8 Sub-network for Osprey problem (agreement).

したがって、事故率だけ見れば現状より安全になるのにもかかわらず、マスメディアでは両機種を同一に扱い、「CV-22」の事故率を持ち出して「オスプレイは危険である」という主張を繰り返しているようにも伺える。なお、著者はネットワーク構築前より本件を知っており、ネットワークを精査することで初めて知り得た事実ではない。次節にて、注目すべきポイントを自動的に抽出する機能について説明する。これにより



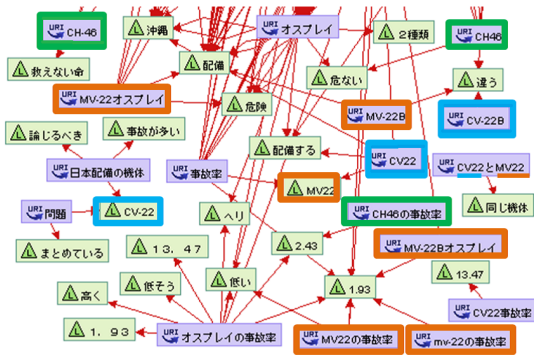


図 9 オspreyに関するサブネットワーク (MV-22, CV-22)

Fig. 9 Sub-network for Osprey problem (MV-22, CV-22).

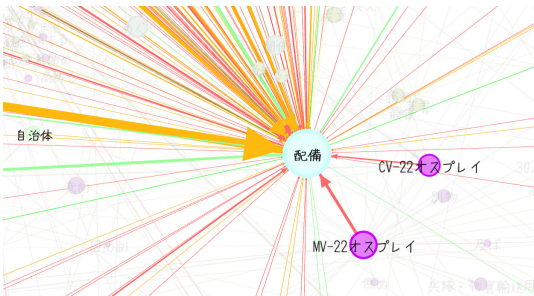


図 10 オspreyに関するサブネットワーク (配備ノードを Activity 属性で参照している様子。ノードの色がソーシャルメディアでしか出現しないことを示す)

Fig. 10 Sub-network for Osprey problem (coloring comparison).

ユーザは一定量の結果リストに目を通すことで有用な知見を得ることができるようになるだろう。

### 3.1.3 事象の偏在性

一般にマスメディアは大衆に向けた粒度の大きい情報を扱うのに対し、ソーシャルメディアでは情報を発信する目的・対象が多様であることから粒度の小さい情報が得られやすいという特徴がある。その一例として地域ごとの情報の偏在性が挙げられる。オspreyの事例では、配備後の影響度合いからか、人々の関心は居住地域と強い相関があった。オspreyに関するツイートは、Location 属性が沖縄と判定されたものが東京と判定されたものよりも多かったことから、ツイート数が人口と比例（あるいは二次曲線的相関）があると想定すると、圧倒的な比率差があるといえるだろう。そこで、地域や距離を考慮したネットワークを

構築することで、その地域特有の問題や現象に対する理解を容易にすることができるだろう。例えば、LODの活用事例でよく見られるように地図情報と連携させることで、特定の事象への賛成反対意見などの影響範囲を可視化でき、マスメディアと対比して注目すべき情報を見つけることができるだろう。

### 3.1.4 事象間の因果関係

更に、ある事象が起こる原因を事象間の関係を辿ることで見つけることができる。オspreyの事例では、先に述べたようにオsprey配備に賛成している意見も見られるが、その理由としては米軍が撤退することになれば、中国の脅威が増すとの考えが見られる。更に、中国脅威論の背景には、昨今の尖閣諸島問題に加えて、中国資本が国内の不動産購入を急速に進めていることなどが伺える（ツイートの例：「沖縄でのオsprey反対者は中国の手先、オsprey配備の目的は尖閣防衛だから」、「沖縄人が全員反対しているかのよう捏造」、「企業だけでなく不動産も中国資本が買い漁っている」など）。このようにネットワークから特定の事象に対する原因（や更にその原因）の可能性を探ることができれば、マスメディアと対比して各主張の裏にある情報の発見を可能とすることを今後検討したい。

### 3.2 比較観点に基づく注目ポイントの抽出

本節では、前節の予備実験（及び、これまでネットワーク構築にあたった経験）から、前節で挙げた多様性、希少性、偏在性、因果関係の4点をメディア比較における観点として設定した。そして、前節とは異なるデータセットから事象ネットワークを構築し、これら4点に基いて注目すべきポイント（ノードとエッジの組合せ）を自動的に抽出、それらをユーザに提示することで本研究の目的であるメディア比較支援の有効性を評価する。次章でも述べるように、本研究の特徴はTwitterなどの自然文からキーワード（インスタンス）をキーワード間の関係を表す事象属性（プロパティ）で繋いだグラフ構造（Linked Data）に変換している点であり、特に本論文では「事象」という単位（一つ以上のトリプルからなる部分グラフ）で一定程度の抽出精度を実現している。そこで、これら事象をSPARQLクエリを用いてグラフ検索することで、上記の観点に沿ったトリプルの組み（事象）を抽出する。ここでは「第46回衆議院議員総選挙・2012年東京都知事選挙」に関するデータセット（収集機関2012/12/04～2012/12/19、約14万ツイート、1220記事）を対象と

表 10 総選挙に関する事象ネットワークに関するデータ  
Table 10 Network information for election.

トリプル数			
ソーシャル	マス	合計	
104,863	105,058	209,921	

ノード数			
ソーシャル	マス	共通	合計
71,016 (62.4%)	32,568 (28.6%)	10,272 (9.0%)	113,856

した. 表 10 にネットワークのノード構成を示す.

**3.2.1 SPARQL クエリによる注目ポイントの検索**  
クエリーコードは紙面の都合から割愛するが, 各観点について以下のようにエッジを辿ることで注目すべきポイントを検索する.

#### 多様性

ソーシャル, マス各データセットを対象に, 両データセット共通ノードの中からエッジの参照元, 参照先という頻度の観点で頻出するノードを絞り込み, そのノードに代表的なエッジとして Activity/Status 属性で繋がる非共通ノードを検索する. 検索された非共通ノード数が多いノードであるほど, 多様な行動・状態・意見が確認できる注目すべきノードとみなせる. 各注目ノードの周辺トリプルを共通行動/状態/意見, 非共通行動/状態/意見 (ソーシャルのみに出現, マスのみに出現) の組みとして抽出する. 注目ノードとともにこれらの情報を提示することで整理された情報をユーザに提示できると考えている.

例えば, 「自民党」ノードに注目した場合, マスメディアは 155 種の Activity, 10 種の Status, 13 種の Regard が接続しているが, ソーシャルメディアでは 2556 種の Activity, 463 種の Status, 379 種の Regard が接続している

#### 希少性

ソーシャル, マス両データセットを対象に, event id を手掛かりに各事象を表す部分グラフを抽出し, グラフの形 (ノード, エッジの組み) が他と最も類似していない事象を抽出する (多様性とは異なり, 両データセットの比較ではない). ただし, ノードの内, 一つは上記, 共通行動/状態/意見を表すトリプルのいずれかに繋がっているものとする (ノイズ除去のため).

例として, 部分グラフ (Subject:公明党, Activity:全勝, Location:小選挙区, Status:凄い) に Cause 属性で繋がる部分グラフ (Subject:死に票, Status:0%) などが挙げられる (組織的に投票されており, 死に票がないの意. ただし, あくまで Twitter 上の一意見で

ある).

#### 偏在性

ソーシャル, マス各データセットを対象に, Location 属性を含むトリプルを検索し, event id を手掛かりに同一事象内の Subject (に対応する文字列), Activity (に対応する文字列), Object (に対応する文字列) を抽出, データセットごとに Location (に対応する文字列) でソートして出力する.

#### 因果関係

ソーシャル, マス両データセットを対象に, Cause/Case/Next 属性を含むトリプルを検索し, Cause 関係にある両事象内の Subject (に対応する文字列), Activity (に対応する文字列), Object (に対応する文字列) を抽出, 2 対 1 組みとして出力する. 例として, 部分グラフ (Subject:初老の男性, Target:安倍総裁, Activity:注意, Location:静岡県内) に Cause 属性で繋がる部分グラフ (Subject:安倍総裁, Time:後から, Activity:座る, Object:席, Location:JR) などが挙げられる (安倍総裁が遊説のため静岡県内を JR 東海道線の普通列車で移動中, 後から乗ってきた安倍氏が座ったために初老の男性に注意された).

#### 3.2.2 ユーザ評価による有効性の確認

SPARQL を用いた重要ポイントの検索手法の有効性を確認するために前節で紹介したクエリによって返却されたトリプルまたは事象をユーザに提示し, 多様性と希少性の観点については, 嗜好性 (興味深い, 面白いなど) の観点で評価した. また, 偏在性と因果関係の観点においては, 嗜好性に加えて未知性 (知らなかった, など) の観点で評価した. 多様性と希少性に関して, 未知性の観点が測られていないのは, この二つは誰もが知っている共通ノード (「自民党」など) を主題としたユーザへの提示となっているため, 知らなかったという未知性の観点と, 知って有用だったという嗜好性の観点をユーザ側で区別しにくいと考えたためである. ユーザによる観点のぶれを防ぐため, 嗜好性として一本化した.

評価では, 四つの比較観点で検索して得られた事象を表す部分グラフ (多様性・希少性: 約 30 件, 偏在性: 約 100 件, 因果関係: 約 50 件) を被験者 5 名 (いずれも大学院生) に紙で提示し, それぞれについて (未知性と) 嗜好性に合うかどうかを 2 択で判断してもらい, 結果を平均した. 表 11 に評価結果を示す. なお, 総選挙の話題は多岐に及び, 再現率を算出するための十分な正解データの作成が困難であったため, 今回は

表 11 注目ポイントの評価  
Table 11 Evaluation of comparable points.

比較観点	多様性	希少性	偏在性	因果関係
未知性に関する適合率 (%)	—	—	51.74%	42.82%
嗜好性に関する適合率 (%)	65.33%	58.33%	55.08%	51.16%

適合率に留めている。

注目ポイントの評価結果から四つの比較観点の嗜好性に関する適合率は全て 50%を超える値を示した。今回、本実験に先立って表 10 にある約 21 万トリプルからランダムに約 100 トリプルを抽出して被験者 3 名に提示、嗜好性を測る予備実験を行ったところ、適合率は平均 14.6%であった。これは、無作為にメディア情報に触れた際に、1～2 割程度、新たな発見があった、という事実と相当しており、日常経験にも沿うものと思われる。この結果から、本実験における 50%超の適合率には一定の有効性があり、四つの比較観点は注目ポイントを推定する上での手がかりになるといえるだろう。また、使用したクエリはコンテキスト情報を用いずにノード・エッジなどのネットワーク構造情報のみを手がかりにしているため、ドメインによらない汎用性があることも確認できた。

なお、被験者は、有権者であり一般常識はもっているが、政治学を専攻するなど特別な識者ではない、というクラスを想定して選出した。データに対する判断力としては、平均的なクラスに入るものと考ええる。また、予備実験の被験者は、同じデータを 2 回見ると結果に影響があると考えたため、同程度の専門性をもつと思われる別の方を選出した。

今後は、更なる精度向上のために、事象抽出精度の改善、クエリの改善、新たな比較観点の追加、各観点をもとにした機械学習手法の適用などが必要だと考えている。

#### 4. 関連研究

本研究と類似した目的から、Web 空間若しくは Twitter から単語間の関連語、話題抽出を行った研究に [11]～[13] が挙げられる。これらは時系列変化における単語の頻度情報、単語間の共起頻度などに着目することで、関連語及び話題抽出が行えることを示した。それに対し、本研究では複数ノード（キーワード）と複数のエッジから事象関係をネットワークで表現しているため、点（単語）同士を頻度でつなげたネット

ワークよりも関係の意味を表現できるようになっており、より深い分析への活用が期待できる。

また、単語・関連語ではなく、語句レベルで語句間の関係を抽出し、効果的な可視化を行うシステムを構築した研究として伊藤らの研究 [14] がある。伊藤らはブログから個人の行動・興味とそれらの対象を係り受け解析を用いて事象として抽出し、入力キーワードに対して関連する事象群の月ごとの変化及び各事象の頻度変化を探索可能にする 3 次元可視化システムを構築した。伊藤らは事象群と文章構造をキーワードを中心としたツリー表現を用いて可視化を行っているが、事象を構成する各ノード（キーワード）同士をつなぐエッジの種類までは定義しておらず、各ノード間の関係を判断するのに人が目視で可視化された情報を確認し推測する必要がある。また、特定事象にあらかじめ固定して詳細に分析することには長けているが、複数の事象を同時に比較するためには複数のキーワードをあらかじめ用意する必要があるため、分析者の想定を超えた未知な事象間の関係を発見することは難しいと考えている。それに対し、我々は各ノード（キーワード）同士をつなぐエッジに 13 種類の関係（Subject/Activity/Object など）を定義した上で、複数ノードを繋いで一つの事象を表現している。これにより、エッジの意味の明確化だけではなく、各事象を表すネットワークが密に繋がる構造となっており、従来よりも未知性ある知見の獲得が可能になると考えている。

更に、マイクロブログに投稿された情報をソーシャルセンサーとみなした研究もある。代表的な研究として、榊らによる地震検知への応用 [15]、荒牧らによるインフルエンザの流行度合いの見積りへの応用 [16], [17] などが挙げられる。榊らは Twitter のツイートから地震関連の語が含まれているツイートに対して、幾つかの特徴を定義し SVM を用いてツイートがターゲットイベントであるかを判別し、実際にターゲットイベントが検出及び発生場所の特定を行う手法を提案した。また、実際に高精度で地震を検知し、ユーザに地震が到達する前にメールで通知するシステムを開発した [15]。荒牧らはインフルエンザに関連の単語が含まれているツイートに対して、投稿者がインフルエンザ患者なのかどうかをインフルエンザ関連語を特徴語とみなした SVM 学習により現実的な精度で判別できることを示した。また、そのモデルを用いてインフルエンザ患者数を見積り、実際のインフルエンザ患者

数と高い正の相関があることを示し、Twitter からインフルエンザの流行を見積もれることも示している。我々の研究も、Twitter から事象群の抽出を行っているため、Twitter をソーシャルセンサーとしてみなした研究とも位置付けられる。しかし、上記の研究ではツイートの単語数、キーワードが何番目の単語か、ツイート中に出現する各単語、各指定キーワードの前後の単語などを特徴とし分類器に学習させているのに対し、我々の研究では、事象を意味ネットワークで表現するなどよりセマンティックアプローチを取っている点が異なる。

## 5. む す び

本論文では、ユーザがメディア情報を多角的な観点から比較することを支援するため、ユーザに代わってソーシャル、マス両メディアから情報を抽出、見える化し、比較ポイントを提示するエージェントサービスの実現を目指し、CRF とヒューリスティックルールを用いた事象ネットワーク構築手法を提案し、精度を評価を行うとともに、事例を通じて有効性を確認した。

今後の課題としては、事象抽出精度、比較観点の抽出精度の改善に加えて、可視化ツールと比較ポイント提示機能との連携が挙げられる。ネットワークを可視化させた状態で、注目すべき点をポップアップ、ハイライトなどで注意喚起できれば利便性を高められるものと考えている。また、本手法はメディア比較にとどまらず、マーケティングや株価予測など事象分析が有効な用途に広く使える技術であると考えており、今後、他ドメインへの適用も検討していきたい。

## 文 献

- [1] T.B. Lee, "Design issues: Linked data," <http://www.w3.org/DesignIssues/LinkedData.html>, 2006.
- [2] C. Bizer, T. Heath, K. Idehen, and T.B. Lee, "Linked data on the Web (LDOW2008)," Proc. World Wide Web Conference, 2008.
- [3] 武田英明, "Linked Data の動向," カレントアウェアネス, no.308, pp.8–11, 2011.
- [4] T.M. Nguyen, T. Kawamura, Y. Tahara, and A. Ohsuga, "Self-supervised capturing of users' activities from weblogs," Int. J. Intelligent Information and Database Systems (IJIIDS), vol.6, no.1, pp.61–76, 2012.
- [5] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proc. International Conference on Machine Learning (ICML), pp.282–289, 2001.
- [6] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," Proc. 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, pp.134–141, 2003.
- [7] A. McCallum and W. Li, "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons," Proc. 7th conference on Natural language learning, pp.188–191, 2003.
- [8] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to Japanese morphological analysis," Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.230–237, 2004.
- [9] K. Radinsky, S. Davidovich, and S. Markovitch, "Learning causality for news events prediction," Proc. World Wide Web Conference, pp.909–918, 2012.
- [10] 越川兼地, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦, "CRF を用いたメディア情報の抽出と LinkedData 化—ソーシャルメディアとマスメディアの比較事例," 合同エージェントワークショップ&シンポジウム (JAWS 2012), 2012.
- [11] A. Marcus, M.S. Bernstein, O. Badar, D.R. Karger, S. Madden, and R.C. Miller, "Twitinfo: aggregating and visualizing microblogs for event exploration," Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2011), pp.227–236, 2011.
- [12] 藤川智英, 鍛冶伸裕, 吉永直樹, 喜連川優, "マイクロブログ上の話題抽出とユーザの態度の分類に基づく流言検出支援システム," 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), 2012.
- [13] 風間一洋, 鳥海不二夫, 榊 剛史, 篠田孝祐, 栗原 聡, 野田五十樹, "東日本大震災時の Twitter データを用いた単語間の関係の時系列変化の分析," 第 26 回人工知能学会全国大会予稿集, 2012.
- [14] 伊藤正彦, 吉永直樹, 豊田正史, 喜連川優, "係り受け解析を用いたブログユーザの行動・興味に関する時系列推移 3 次元可視化システム," 信学論 (D), vol.J95-D, no.7, pp.1454–1466, July 2012.
- [15] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol.25, no.4, pp.919–931, 2012.
- [16] E. Aramaki, S. Maskawa, and M. Morita, "Influenza patients are invisible in the web: Traditional model still improves the state of the art web based influenza surveillance," Proc. AAAI Spring Symposium, 2012.
- [17] E. Aramaki, S. Maskawa, and M. Morita, "Twitter catches the flu: Detecting influenza epidemics using Twitter," Proc. 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), pp.1568–1576, 2011.

(平成 25 年 2 月 10 日受付, 6 月 21 日再受付)





川村 隆浩

1994 年早稲田大学大学院理工学研究科電気工学専攻修士課程了。同年、(株)東芝入社。現在、同社研究開発センター主任研究員。工学博士（早稲田大学）。2001–2002 年米国カーネギー・メロン大学ロボット工学研究所客員研究員。2003 年より電気通信大学大学院情報システム学研究科客員准教授。2007 年より大阪大学大学院工学研究科非常勤講師。主としてマルチエージェントシステム、セマンティック Web の研究・開発に従事。人工知能学会、情報処理学会各会員。



越川 兼地

2013 年電気通信大学大学院情報システム学研究科社会知能情報学修士課程了。同年、株式会社リクルートホールディングス入社。現在に至る。



中川 博之（正員）

1997 年大阪大学基礎工学部情報工学科卒。同年、鹿島建設（株）入社。2007 年東京大学大学院情報理工学系研究科修士課程了。2008 年同大学院博士課程中退。同年より電気通信大学助教。現在に至る。エージェント及び自己適応システム開発手法の研究に従事。情報処理学会、電子情報通信学会、IEEE CS 各会員。



清 雄一

2009 年東京大学大学院情報理工学系研究科博士後期課程了。同年、(株)三菱総合研究所入社。同社情報技術研究センター金融ソリューション本部等に所属。2013 年より電気通信大学助教。現在に至る。分散コンピューティング、セキュリティ、プライバシー保護技術等の研究に従事。IEEE CS 会員。



田原 康之

1991 年東京大学大学院理学系研究科数学専攻修士課程了。同年、(株)東芝入社。1993–1996 年情報処理振興事業協会出向。1996–1997 年英国 City 大学客員研究員。1997–1998 年英国 Imperial College 客員研究員。2003 年国立情報学研究所入所。2008 年より電気通信大学准教授。博士（情報科学）（早稲田大学）。エージェント技術、及びソフトウェア工学などの研究に従事。情報処理学会、日本ソフトウェア科学会各会員。



大須賀昭彦（正員）

1981 年上智大学理工学部数学科卒。同年、(株)東芝入社。同社研究開発センターソフトウェア技術センター等に所属。1985–1989 年（財）新世代コンピュータ技術開発機構（ICOT）出向。2007 年より電気通信大学大学院情報システム学研究科教授。2012 年より国立情報学研究所客員教授兼任。工学博士（早稲田大学）。主としてソフトウェアのためのフォーマルメソッド、エージェント技術の研究に従事。1986 年度情報処理学会論文賞受賞。IEEE Computer Society Japan Chapter Chair。人工知能学会理事、日本ソフトウェア科学会理事を歴任。情報処理学会、電子情報通信学会、人工知能学会、日本ソフトウェア科学会、IEEE CS 各会員。