画像照合に基づく多メディア間の話題追跡システム

豊田 正史　伊藤 正彦 *　Cai-Zhi Zhu　佐藤 真一　宇野 毅明 †　喜連川 優 ‡

概要.　本論文では，Web と放送映像との間の画像の対応関係に基づき話題の追跡を行える新たなシステムを提案する．社会事象を解析する際には，話題がどのメディアから始まったかを同定し，その話題がどのようにメディア間に広がっていったかを分析する必要がある．Web と放送映像との対応を取るため，Web 上のテキストと放送の字幕情報を用いることも可能であるが，字幕の量は十分ではなく，字幕が付けられていない番組も多数存在するため，映像情報の対応関係を解析する事が不可欠となる．提案するシステムでは，ブログ等の Web メディアに投稿された大量の画像から大きな話題になり多数出現した類似画像を抽出し，これが放送映像を起源とするものかどうかを放送映像アーカイブと照合して，話題の時差や露出の差を調査できる手法を実現した．本システムを用いて，大規模なブログアーカイブと放送映像アーカイブを分析した事例を幾つか示すことで本システムの有用性を示す.

1　はじめに

Web はその出現から約 20 年の間に，ニュースサイト，掲示板，ブログ，ソーシャルメディア，コンテンツ共有サイトなど多数のメディアの集合体として進化を続けてきた．現在は，これら Web 上のメディアとテレビや新聞などの従来型マスメディアを包含する**多メディア**が人々の社会活動を反映し，かつ影響を与えていると言える．こうした状況において，社会事象を分析する際には話題がどのメディアから出現したかを同定し，その話題がどのように多メディアの間に広がっていったかを分析する必要がある．現在においても Web メディアはマスメディア，特にテレビ放送から受ける影響が大きく，Web メディアとテレビ放送の間の話題追跡は重要な課題である.

従来，Web メディアとテレビ放送での話題分析においては，Web 上のブログやソーシャルメディアなどの記事および番組表や字幕など，テキスト情報を用いたものが主であった．しかし，これらのメディアでは積極的に映像・画像を用いることで，文章のみでは伝えることが困難なイベントや事件事故などの様子を伝えており，また字幕の量は十分ではなく，字幕が付けられていない番組も多く存在するため，映像情報の対応関係を解析する事が不可欠となる.

Web 上で話題になった画像が，放送映像中では同じように現れているのか，また，その出現に前後関係はあるのか，話題や画像ごとに出現の仕方に違いや類似性があるのか，逆に特徴のある画像群は具体的にどのような話題を扱っているのか等を調査することにより，以下のような社会事象や多メディア間で及ぼされる影響の分析が可能になる.

- ニュースが各メディアにおいてどのような画像で扱われていたか，メディアにより現れる画像に違いがあるかを調査する.
- イベントや集会などの模様や賑わいを視覚的に把握することで，メディアによる扱われ方の差があるかどうかを調査する.
- 商品画像の出現数の変化から，人気の度合いやデザインの変化を把握する.

本論文では，Web メディアに投稿された大量の画像から大きな話題になり多数出現した類似画像を抽出し，これが放送映像を起源とするものかどうかを放送映像アーカイブを照合して抽出し可視化することで，話題の時差や露出の差を調査できる手法を提案し，それを実現するシステムを実装する．本システムを，大規模なブログアーカイブと放送映像アーカイブに適用し，実際に観測された社会事象の事例を示すことで本システムの有用性を示す.

以下，第 2 節では関連研究について述べ，第 3 節で提案システムの詳細を説明する．第 4 節において本システムを用いた実験結果を示し，第 5 節で本論文をまとめる.

2　関連研究

従来，Web メディアおよびテレビ放送における話題分析においては，単独のメディアを用いたものが多く，複数メディアを扱ったものでも Web 上のブログやソーシャルメディアなどの記事，および番組表や字幕などテキスト情報を用いたものが主であり，大規模な画像照合を用いて多メディア間の情報追跡を行えているものは筆者の知る限りでは存在しない.

Copyright is held by the author(s).

* 東京大学生産技術研究所
† 国立情報学研究所
‡ 東京大学生産技術研究所・国立情報学研究所

Adar らは，検索エンジンの検索クエリとテレビ番組情報の出現時差の解析を行っており [1]，Yang らは，ブログ，新聞，テレビ等の複数メディアにおける話題の時系列パタンを抽出しているが [7]，これらは基本的にテキストデータに現れる語やフレーズの出現頻度を用いたもので画像情報は用いられていない．Luo らは，大規模なニュース画像アーカイブの可視化，および話題分析を行っているが [5]，複数のメディアにまたがる分析までは行えていない．

3 多メディア間話題追跡システム

本論文で提案するシステムは，大規模な Web メディアのアーカイブを用いてユーザの指定した調査対象に関連する記事および画像を抽出し，さらに放送映像のアーカイブを用いて画像照合を行い，抽出した Web 画像に対応する放送映像での露出を抽出することで，これら画像の頻度，時間推移，および出現時差を調査可能にするものである．本システムは，Web メディアアーカイブとして，ブログやソーシャルメディアのように，各記事に対して書き手と書かれた時間が紐付けされた文書アーカイブを想定しており，全文検索インデックスによって記事が検索可能となっているものとする．また，放送映像アーカイブに関しては，画像を入力としてその画像に類似する映像が放送されたショットを検索可能とする検索インデックスを実現して用いている．

ユーザは，調査対象に関する複数の検索語を入力としてシステムに与え，システムは，以下のような手法で，画像抽出，分析，可視化を行う．

1. Web メディアアーカイブを検索し，与えられた検索語を含む記事の集合を取得する．各記事はそれが書かれた時間，書き手，および記事本文を含む．
2. 取得した記事の本文に含まれている画像をすべて取得し，各画像の周辺のテキストを抽出する．
3. Web メディア上で話題となり数多くコピーされた画像を抽出するため，ほぼ同一とみなせる画像のペアを列挙しクラスタリングを行う．この結果得られる *Web* **画像集合**は単独の画像が多数出現したことを表し，画像に基づく話題を構成する要素となる．
4. 抽出した各 Web 画像集合を検索クエリとして用いて，放送映像アーカイブを検索し，類似画像が放送されたショットの集合を抽出する．抽出された各ショットの集合を，便宜上 Web 画像集合に対応する放送画像集合と呼ぶ．
5. Web 画像集合を，それらの周辺テキストの類似度を用いてクラスタリングすることで，高頻度で出現する画像の類似集合である *Web* **画像トピック**を抽出する．さらに Web 画像トピックに含まれる Web 画像集合群に対応する放送画像集合を集約することで，**放送映像トピック**を抽出する．こうして得られた Web 画像トピックとそれに対応する放送映像トピックの時間推移を 3 次元空間に可視化し，メディア間におけるトピックの推移およびその差を閲覧探索可能とする．

以下では，ほぼ同一画像のクラスタリング，放送映像検索の実現方法，トピックの 3 次元可視化手法についてより詳細に説明する．

3.1 ほぼ同一画像のクラスタリング

Web 上の画像は，オリジナルの画像に対し拡大，縮小，回転や，切り取り，一部変更などの加工を経て公開されるものが多く，ほぼ同一の画像をクラスタリングするための類似度算出にはこうした加工に対し頑健な特徴量を用いる必要がある．本手法ではこうした状況に適した局所特徴量の 1 つである SIFT 特徴量 [4] を用いている．SIFT を用いた画像類似度算出では，各画像から抽出された多数の SIFT 特徴点の類似度を算出し，画像間でマッチする特徴点の数をカウントする必要がある．画像数は通常多数に及ぶため類似する SIFT 特徴点のペアを高速に列挙する手法を開発した．本列挙手法は，宇野による大規模部分文字列マッチングアルゴリズム [6] を基にノイズとなる特徴点の除去処理を加えて改良を施したものである．

各画像からは，サイズの正規化を行った上で，最大 100 個の SIFT 特徴点を抽出する [1]．SIFT 特徴点は，128 次元の数値ベクトルで表わされるが，各値を中央値以上か，それ以下かの 2 値に量子化し，128 ビットのベクトルに変換する．さらにこの数値ベクトルは 8 個の値からなる 16 のブロックから構成されることを考慮し，各 8 ビットを 1 文字とみなすことで書く特徴点を 16 文字の文字列と考える．宇野の類似文字列列挙手法は，大量の固定長文字列から，ハミング距離が一定以下の文字列ペアを高速に列挙するもので，本手法ではほぼ同一の画像を抽出するため，ハミング距離 1 以下の類似特徴点のペアをマッチしたものと見做して列挙し，画像間でマッチする特徴点の数を，画像ペアをキーとしたハッシュ表を用いて算出する．2 つの画像の類似度は，画像間でマッチする特徴点の数で表わされる．

SIFT 特徴点を用いた類似画像抽出においては，非常に多数の特徴点とマッチする特徴点が，ノイズとして類似度の精度に悪影響を与えることが分かっ

[1] ここでは特徴点の最大数を経験的に 100 個に定めたため，これが最適であるとは限らない．しかし最大数をこれ以上大きくすると，ブログ用の縮小画像や携帯電話などで撮影した小さな画像に関して特徴点の数が多くなりすぎノイズが増すことが実験から分かっている．

ており，本手法では，類似ペア列挙の途中段階で，閾値以上の特徴点とマッチする特徴点をフィルタする処理を加えることで，ノイズに対して頑健な類似画像抽出を実現している．ノイズとなる特徴点の数は非常に少ないため，計算量を大きく削減するものではないが，大量の類似特徴点のペアを出力した後フィルタする必要がなく直接類似画像ペアを出力できるため，中間出力を削減することができる．本手法は，4 百万画像から抽出した 4 億の SIFT 特徴点の集合を用いて類似画像ペア (閾値 1000 の場合，約 1.35 億ペア) を列挙するのに，10GB のメモリと 20 コアを用いて 2 時間程度しかかからないスケーラビリティを持っている．

類似する画像のペアが列挙できると，画像をノード，類似度をエッジとしたグラフを作ることが可能となる．本手法では，このグラフに対して，スタークラスタリング [2] という簡易なクラスタリング手法を適用することで，画像集合を抽出している．スタークラスタリングは，次数の一番高いノードを中心とし，閾値以上の類似度を持つ隣接ノードを，中心と同じクラスタとして出力するという処理を繰り返す単純なものである．本手法では，あるノードについて後からより類似度の高い中心が見つかった場合，元の中心からその中心へ所属を切り替えるという簡易な改良を施して用いている．スタークラスタリングの計算時間はノード数とエッジ数の和に比例しており，大規模なデータをインタラクティブにクラスタリングするのに適している．また，類似度の低いノードをあえてクラスタに所属させることは無いため，ノイズの多いデータのクラスタリングにも適している．こうした理由から本手法ではスタークラスタリングを用いているが，他のクラスタリング手法を用いても良い．

3.2 放送映像の検索手法

放送映像アーカイブ中のショットを任意の画像により問い合わせ可能な，問い合わせ画像中と同一の物体を含む映像を検索可能なシステムを [8] の手法に基づき構築した．まずオフライン処理として，放送映像アーカイブを構成する各映像クリップから，非常に密にフレーム画像を切り出し（例えば 3 フレームごとに 1 フレーム），各フレーム画像から密に SIFT 等の局所特徴量を抽出する（たとえば各フレーム画像から数百点，SIFT もしくは color SIFT を使用）．局所特徴量は階層的 k-means により，100 万コードブックというきわめて疎な離散化を行い，各クリップにつき一つの Bag of Feature (BoF) ベクトルを得る．次いで，検索時のオンライン処理では，与えられた問い合わせ画像から，同様に BoF ベクトルを得る．これとデータベース中の BoF ベクトルとの間で L1 距離を求め，ランキングリストを得ることにより，検索を実現している．検索は転置索引により効果的に高速化可能である．

この方法で実際に検索を行うと，大規模なデータベースが対象となった場合，検索結果にノイズがきわめて多くなってしまうことが判明した．そこで，幾何整合性判定によりリランキングを行う技術 [3] を導入して，精度の向上を図っている．

3.3 多メディア間のトピック可視化・閲覧手法

本システムでは，Web メディアおよび放送映像から抽出されたトピックの時系列推移を可視化することで，多メディア間の影響分析を視覚的に可能にする新たな可視化手法を提案する．

図 1 に本可視化手法の概要を示す．平面上に，時間軸とトピックの軸を持ち，各時間軸上に Web メディアアーカイブから抽出された Web 画像集合，および放送映像アーカイブから抽出した対応する放送画像集合をヒストグラムとして可視化している．画像は単位時間ごと（例えば 1 日ごと）に集約され，その時間に出現した画像を高さ方向に積み上げることで，画像による棒グラフから構成されるヒストグラムとして表現される．

画像の出現の時間差を認識しやすくするため，Web メディアと放送映像の画像ヒストグラムは，図 1 のように並列して配置することもでき，また，図 3 のように上下にヒストグラムを反転させて配置することも可能である．

上記の手法で抽出された Web 画像集合は，さらにその周辺テキストを用いて意味的にトピックにまとめることも可能である．図 1 では，各 Web 画像集合に含まれる画像の周辺テキストをまとめて 1 文書とみなし，tf-idf で重み付けした単語頻度ベクトルを算出し，コサイン類似度を用いて上記のスタークラスタリングを行い，*Web* **画像トピック**にまとめている．Web 画像集合のクラスタリングに関しては，スタークラスタリングの閾値をインタラクティブに変更しながらクラスタリング結果を閲覧するシステムを用意し，適切な閾値を選択可能としている (図 2)．こうして得られた Web 画像トピックに対し，放送映像側には各 Web 画像集合に対応する放送画像集合をまとめて**放送画像トピック**として表示している．

ユーザは，Web 画像数およびショット数の時系列に対する交差相関係数，コサイン類似度等のレンジをパラレルコーディネートを用いたダイナミッククエリにより対話的に指定することで，様々な時間差をもつ画像クラスタを抽出し，容易に詳細を探索することが可能である．交差相関係数を用いると，指定したメディアで先行して（もしくは遅れて）現れた画像集合を抽出することができ，放送先行もしくは Web メディア先行で流行した話題を抽出することが可能になる．時系列間のコサイン類似度が高い画像集合を抽出することで出現タイミングが似てい

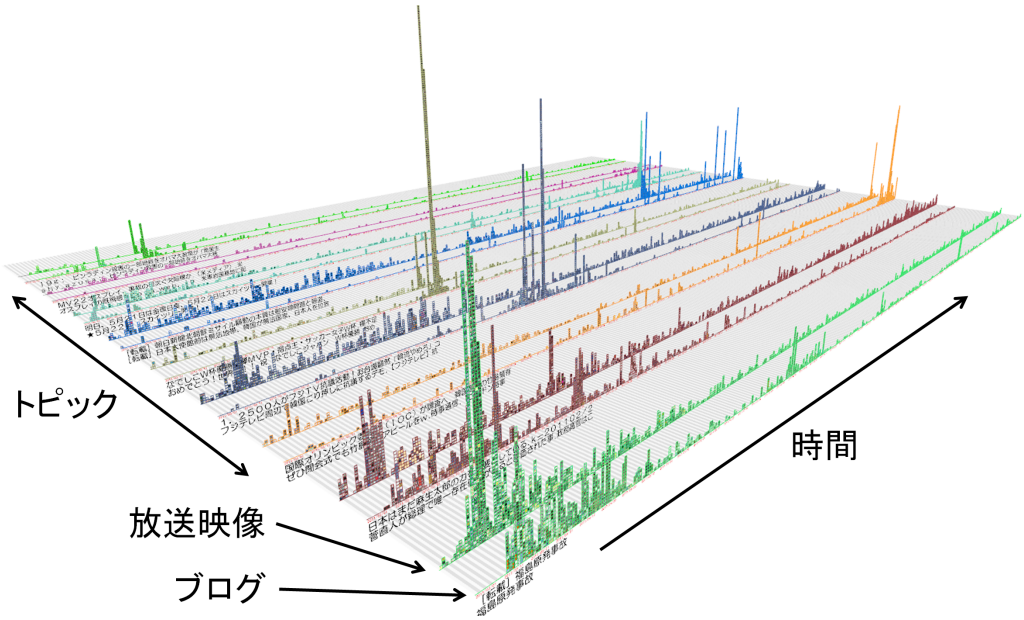

図 1. 複数トピックの推移の概要を可視化

る話題を抽出でき，類似度が低い画像集合からは，放送もしくは Web メディアのみで流行した話題を抽出することができる．

4 大規模ブログアーカイブおよび大規模放送映像アーカイブを用いた実験

4.1 データセット

前節のシステムを実装し，我々が東京大学生産技術研究所において収集している大規模ブログアーカイブおよび大規模放送映像アーカイブを用いた実験を行った．Web メディアとしては，我々が 2006 年から収集している延べ 200 万ブログを対象とした RSS/Atom フィードのアーカイブを用いた．特に，社会分析のターゲットとして 2011 年 3 月から 2012 年 9 月に起きた，東日本大震災およびそれに引き続く原発事故，サッカー女子ワールドカップ，北朝鮮ミサイル問題，ロンドンオリンピックなどを想定して，検索用のキーワードを約 90 個程度選出し，アーカイブから抽出したブログ記事に含まれる約 400 万の画像を実験データとして用いた．

また，国立情報学研究所においてアーカイブした，2011 年 3 月より 2012 年 9 月までに放送された全チャンネルのニュース番組，合計約 6000 時間分を対象として検索インデックスを構築した．検索インデックス構築対象の 6000 時間分のニュースに対し，連続フレーム間の色ヒストグラム差分に基づくショット分割を行い，約 300 万ショットを検出した．これが検索対象の単位となる．さらに，各フレーム単位での詳細な物体検索を実現するため，1 秒おきに約 2000 万キーフレームを抽出した．局所特徴量数は約 200 億となる．こうして構築したデータベースに対し，ブログから想定したターゲットに関連するキーワードに基づいて抽出した 4000 の Web 画像集合 (46000 画像) を問い合わせとして照合を行い，相互の影響解析につながる根拠となるような画像が数多く検出されたことを確認した．

4.2 事例

本節では上述のデータセットを用い，提案システムを用いて抽出した多メディア間の話題追跡の事例を示す．

図 1 は，震災，原発，オリンピック等の検索語でブログアーカイブから抽出した Web 画像トピック，さらに放送映像アーカイブから抽出した対応する放送画像トピック，それぞれをトピック毎，一日毎に集約し時間軸上に可視化している．全体像を俯瞰（図 (a)）することで，トピック間の類似性や，ブログ・放送映像間で同時にバーストしている期間などを観測できる．さらに，時系列に対する交差相関係数，コサイン類似度等から様々な反応差をもつ画像集合を対話的に抽出し詳細を探索することが可能である．

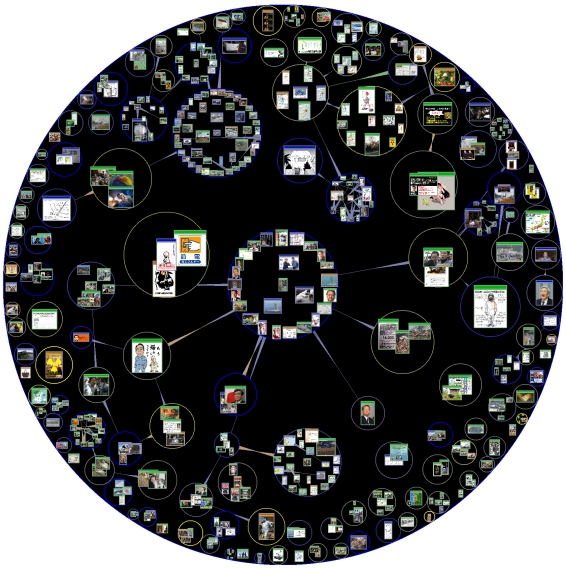

図 2. Web 画像集合のクラスタリング結果の可視化

この中から，東日本大震災の際の福島第一原発事故に関するトピックに着目し，拡大したものが図 3 である．このトピックに含まれる画像の多くは，事故のあった建屋を撮影したもので，画像ヒストグラムの形状，及び写真の出現順序を見ることで，これらがニュース番組において繰り返し使用され，その後ブログ上でも広まったことが見て取れる．同種の写真は，放送のみならず新聞等のメディアにおいても使用されており，この場合においてはマスメディアが先行して写真を掲示し，それがブログに広まったと解釈することができる．

本システムのダイナミッククエリ機能を用いると，このトピックの中からブログが先行して用いた画像があるかどうかを調べることができる．図 4 は，交差相関係数を用いてブログが先行している画像クラスタを抽出したものである．この画像は，MIT の研究者によるブログにおいて福島第一原発事故解説に用いられた原子炉の透視図である．本記事が日本のブログにおいて翻訳され話題となり，その後 2 日遅れてニュース番組で用いられるようになったことが分かる．番組作成者が実際にブログを見てこの画像を発見したかどうかまでは不明であるが，これは画像の照合を用いることで初めて発見できる映像素材の原典追跡の事例となっている．メディア社会学と呼ばれる学問分野においては，映像の原典を探ることが重要なテーマの一つとなっており，本手法はそうした調査に有用である．

テレビとブログとでは，同じ話題に対する反応の大きさや勢いが異なり，それは国際問題，政治問題において顕著であることが今回の実験結果から分かっている．図 5 はその一例を示したものである．2012 年のオリンピックにおいて，試合の終了時にある選

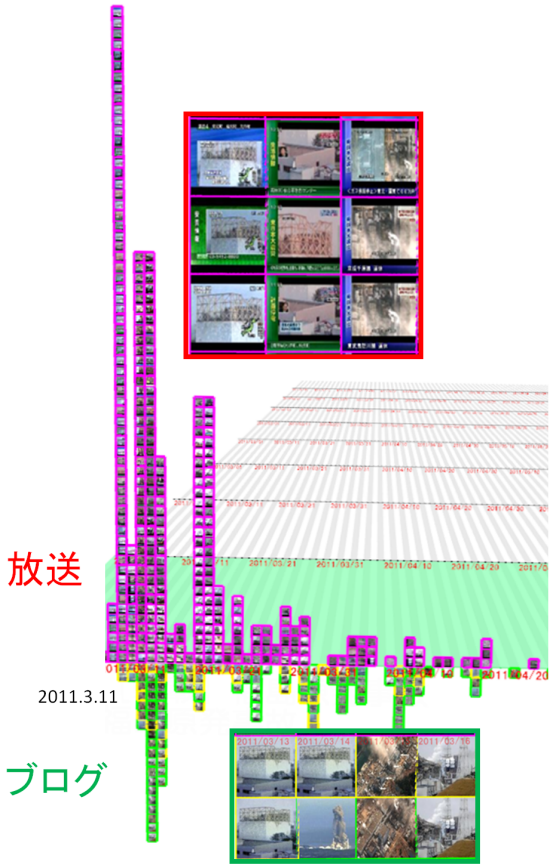

図 3. 福島第一原発事故に関するトピックのメディア比較

手が政治的なアピールをしたことが問題となった．そのニュースは最初テレビで放送されたが，その画像の露出は少なかった．しかし，このニュースは多くのブロガーを刺激しブログ上ではこの画像が大量に使われることとなった．この大きな反応を受けて，テレビのニュースでもこの問題が大きく扱われるようになった．こうした，テレビとブログが相互に反応して話題が大きくなるケースは，センシティブな国際問題や政治問題において良く見られるパターンである．

5 まとめ

本論文では，ブログ等の Web メディアに投稿された大量の画像から大きな話題になり多数出現した類似画像を抽出し，これが放送映像を起源とするものかどうかを放送映像アーカイブを照合して，話題の時差や露出の差を調査できるシステムを実現した．さらに大規模なブログアーカイブと，放送映像アーカイブを用いて実験を行い，様々な社会現象に関する多メディア間の情報追跡を行えることを示した．

本システムを用いることで，過去に起きた様々なイベントに対しての Web メディアと放送映像での反応の違いを学ぶことが可能であり，得られた知見

図 4. ブログがテレビ放送に先行した画像の例

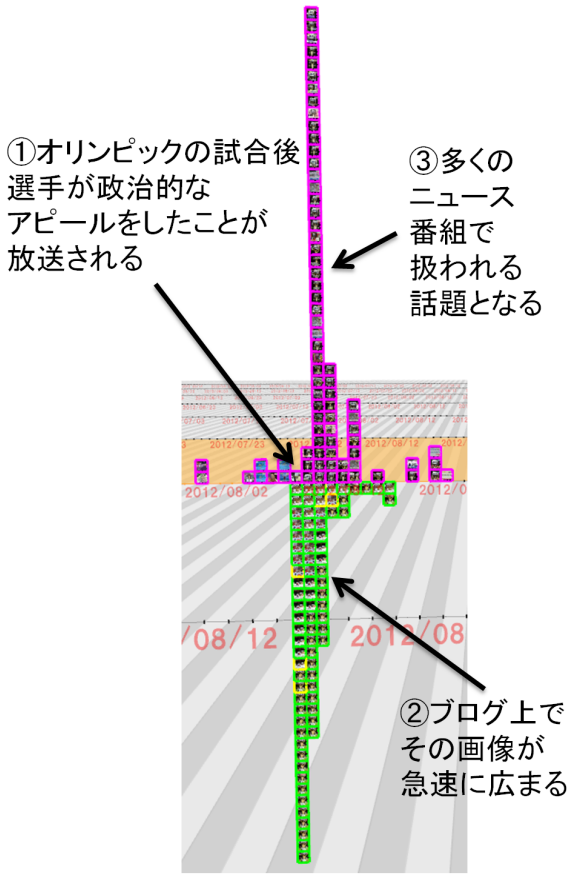

図 5. テレビとブログが相互に影響を与えた例

は今後同種のイベントが起きた際にどのような反応が各メディアで起こるかを予想する際の手がかりとなり得る．また，一方のメディアで話題になっているのに，もう一方のメディアでは大きな話題にならないケースも，幾つか本システムで観測されており，その原因を探ることで企業や組織の広報活動に活用することも可能である．ここで示した事例は，本実験で得られた知見のごく一部であるが，本システムを用いることで多メディアに関わる様々な興味深い社会事象を捉えられる可能性を示せていると信じる．

現状では，Web 画像集合と対応する放送画像集合が算出された後はある程度インタラクティブに閲覧・探索が行えるが，そのための事前計算には依然としてコストが高い．画像の特徴量抽出や放送映像データベースの索引を事前に作成しておくことはできるものの，画像類似度の算出や，対応画像の検索はインタラクティブに行えるほどではないため，一層の高速化が必要である．

謝辞

本研究は，文部科学省「多メディアＷｅｂ解析基盤の構築及び社会分析ソフトウェアの開発」，および JSPS 科研費 25280111 の助成を受けたものです．

参考文献

[1] E. Adar, D. S. Weld, B. N. Bershad, and S. D. Gribble. Why We Search: Visualizing and Predicting User Behavior. In *Proceedings of WWW'07*, pp. 161–170, 2007.

[2] J. A. Aslam, E. Pelekhov, and D. Rus. The Star Clustering Algorithm for Static and Dynamic Information Organization. *Journal of Graph Algorithm and Applications*, 8(1):95–129, 2004.

[3] K. Lebeda, J. Matas, and O. Chum. Fixing the Locally Optimized RANSAC. In *Proceedings of BMVC*, 2012.

[4] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[5] H. Luo, J. Fan, J. Yang, W. Ribarsky, and S. Satoh. Analyzing Large-Scale News Video Databases to Support Knowledge Visualization and Intuitive Retrieval. In *Proceedings of VAST'07*, pp. 107–114, 2007.

[6] T. Uno. Multi-sorting algorithm for finding pairs of similar short substrings from large-scale string data. *Knowledge and Information Systems*, 25(2):229–251, 2010.

[7] J. Yang and J. Leskovec. Patterns of Temporal Variation in Online Media. In *Proceedings of WSDM'11*, 2011.

[8] C. Z. Zhu and S. Satoh. Large Vocabulary Quantization for Searching Instances from Videos. In *Proceedings of ICMR'12*, pp. 52:1–52:8, 2012.