



ソーシャルメディア時系列解析のための3次元情報可視化*

伊藤 正彦**

3D Information Visualization for Analyzing Temporal Changes in Social Media Contents

Masahiko ITOH

1. はじめに

ソーシャルメディアにより、人々が自身の興味、どこで何をしたかなどの活動内容、あるいは社会問題や商品に対する主観的意見などを、いつでもどこでも発信できるようになって久しい。文章だけでなく、写真や動画、ユーザ間のコミュニケーション、さらには位置情報までも含めた膨大な量のデータが日々生み出されている。このような、時間・社会状況とともに変化する人々の生の声は、製品、人物、政策等の評判、あるいは、交通システムなど社会インフラに関する意見など、マーケティング、社会分析、都市計画など様々な観点から重要なデータになる。また、ソーシャルメディアの普及は、従来型のマスメディアのあり方も変えつつあり、ソーシャルメディアがマスメディアの情報発信源となる新たな流れも生まれている。マスメディアとソーシャルメディアがお互いに影響を与え合う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、どのように他のメディアに広がったのかを分析することも重要になってくる。

著者らは、膨大かつ多様なソーシャルメディアデータから、流行の変化などを視覚的に探索するための様々なシステムを研究開発してきた。多様なシステムを開発するにあたり、汎用的な部品の組み合わせで時系列情報の3次元可視化システムを実現するフレームワークを提案し、ソーシャルメディアから得られた様々な時系列データに適応することで、可視化・探索システムを構築してきた。フレームワークの詳細は文献¹⁾²⁾を参照されたい。

本稿では、著者らがこれまで構築してきた、ソーシャルメディアのリンク構造、文章構造、類似画像クラスタ、画像出現のメディア間差異、および、位置参照情報に着目した応用システムに関して探索事例を中心に紹介する。

本稿で紹介する事例では、東京大学生産技術研究所の

喜連川・豊田研究室において、2006年から構築している大規模ブロッグアーカイブ、および、2011年から構築している大規模ツイッターアーカイブをソーシャルメディアデータセットとして用いている。また、5章の例では、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6チャンネル分ののニュース番組を放送メディアデータセットとして用いている。

2. 時系列リンク構造可視化によるブログ上のインフルエンサー追跡システム

本章では、話題に関する文書数の推移と同時に、リンク構造の変遷についても把握可能とするWebグラフのインタラクティブな3次元可視化システムを紹介する³⁾。

本システムはブロッグアーカイブから抽出された特定の話題に関するリンク構造を表すWebグラフを3次元空間中の時間軸に沿って可視化する。この例では、ブログのエントリをノード、エントリ間のハイパーリンクをエッジとするネットワークを描画している。

ある時間のWebグラフを表示するパネルをTimeSliceと呼ぶ。ユーザはTimeSliceをマウスでドラッグすることでグラフの変化をアニメーションさせながら任意時間におけるグラフを閲覧できる。TimeSliceの側面には文書数を表すヒストグラムが表示されており、急激な増加など特徴的な変化が起きた時点でのグラフを容易に認識し、表示することが可能になる。また、異なる時間のWebグラフを比較するため、新たなTimeSliceを自由に追加することができる。これにより、ブログ空間上の話題に関する話題の中心、興味の広がりなどの時間変化を探索可能にした。さらに、より詳細な比較を可能にするタイル表示および重畳表示手法を3次元空間で統合し、これらをシームレスに切り替え可能にした。これらにより、ユーザは変化の全体像を俯瞰しながら、より局所的な変化の詳細を観測することが可能になった。

Fig. 1 にブログ空間上における「ワーキングプア」に関連する人々の興味の移り変わりを可視化した例を示す。ワーキングプア問題を広めるきっかけとなったNHKの番組にリンクが集中している様子が見て取れ、第1回、

* 原稿受付 2016年1月15日

** 非会員 東京大学 生産技術研究所
(〒153-8505 東京都目黒区駒場4-6-1,
E-mail: imash@tkl.iis.u-tokyo.ac.jp)
独立行政法人 情報通信研究機構
ソーシャルICT推進研究センター

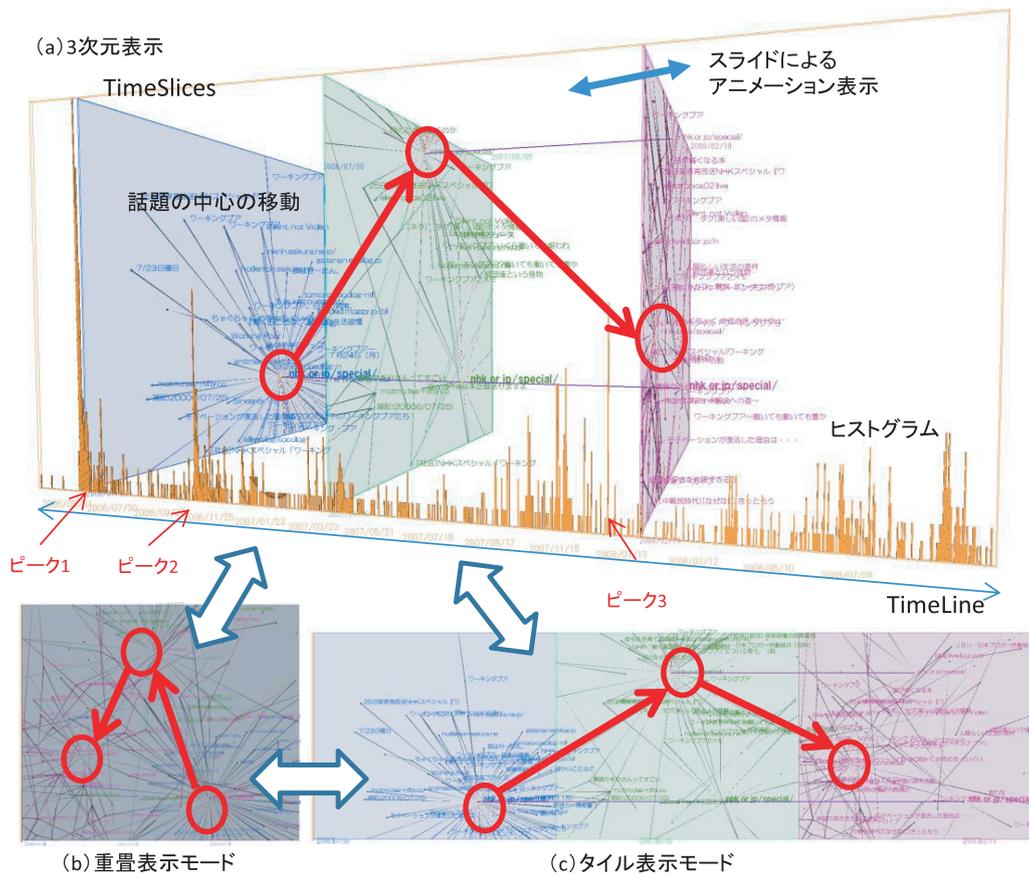


Fig. 1 時系列リンク構造可視化によるブログ上のインフルエンサー追跡.

第2回, さらには第3回の放送に伴って, 話題の中心が移動している様子が観測できる.

3. 係り受け解析を用いたブログ上の話題追跡システム

ウェブ上での人々の興味・行動の変化を分析するために, 多くのサービスが開発されているが, 多くは出現頻度の変化の提示や関連語の提示のみであり, 実際にどのような文脈で話題になっているかを把握することは困難である. 本章で紹介するシステム⁴⁾では, プログアーカイブに対して係り受け解析を行い, 注目する単語と関連語が直接共起する文脈をイベントとして収集し, まとめて可視化することでユーザーに提示する. さらに, イベントの情報を段階的に詳細化しながら探索する仕組みにより, 興味を持ったイベントに関して, 誰が, どこで, 何を, なぜ, などの行動・興味の詳細を調査することが可能になる. これらにより, そのキーワードがどのような文脈で話題になっているかに関して, より容易に理解することが可能となる.

このシステムでは, 与えられたキーワードに関連するイベントとその詳細情報を集約し, それらの構造をツリー表現により可視化し俯瞰可能にする. また, 3次元空間の一軸を時間軸として用いることで, イベント群の構造の時間変化と各イベントの出現頻度の時間変化を俯瞰可能にする.

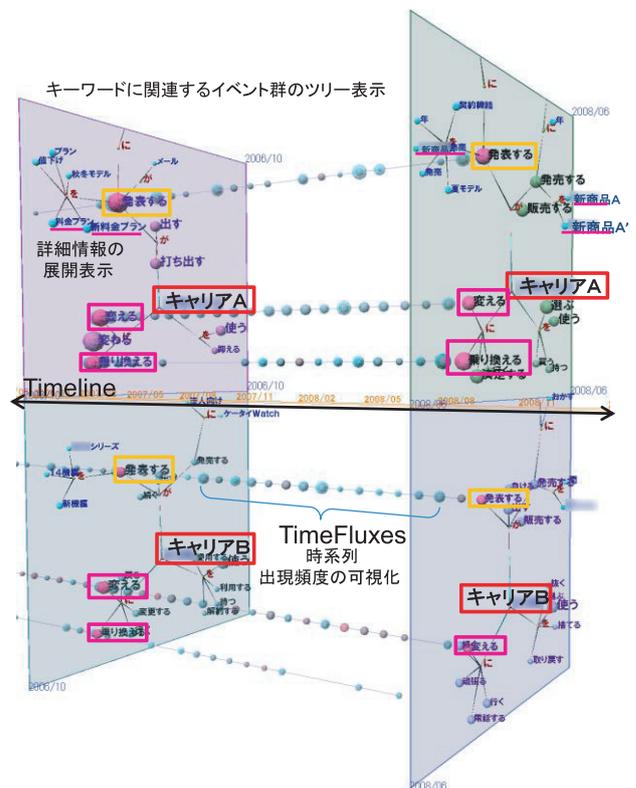
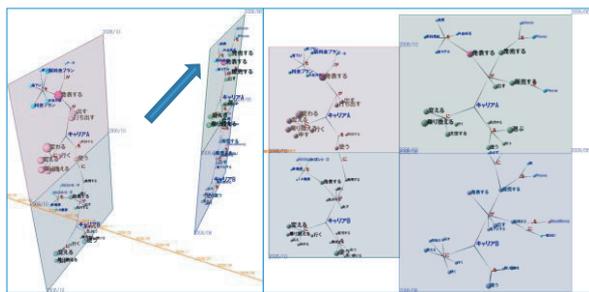
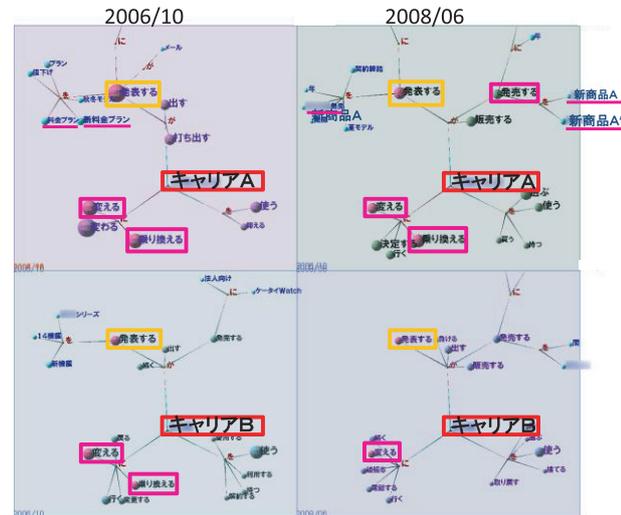


Fig. 2 係り受け解析を用いたブログ上の話題追跡システム.



(a) TimeSliceの奥行き方向へのスライド (b) 視点を時間軸方向へ移動



(c) 正射影によるタイル表示の実現

Fig. 3 シームレスなタイル表示への切り替えインタフェース.

Fig. 2 の例は、複数の携帯キャリアのマーケティング効果を可視化、比較した例である。図において、上部の TimeSlices はキャリア A に関する話題を可視化している、下部の TimeSlices はキャリア B に関する話題を可視化している。キャリア B に比べてキャリア A では「乗り換える」「変える」の記述が目立つ。TimeFluxes を表示すると、キャリア A 側では何度か「乗り換える」ピークが来ていることが確認できる。図における左側の TimeSlices は一度目のピークが来たタイミング（2006年10月）に合わせてある。右側の TimeSlices は二度目のピークが来た 2008年6月に合わせてある。これらのタイミングでキャリア A で何らかの発表もしくは発売をしていることが、ノードのサイズから伺える。「発表する」「発売する」に関して、詳細を展開表示することで、一度目のピークでは「新料金プラン」、二度目のピークでは「新商品 A」を発表している事が分かる。これらからキャリア A のマーケティング戦略は社会に大きなインパクトを与えているであろうことが見て取れる。一方、キャリア B でも何度も「発表する」の頻度が大きくなってきているにもかかわらず、それが「乗り換える」や「変える」という行動にあまり大きな影響を与えていないことが伺える。

Fig. 3 は Fig. 2 をタイル表示に切り替えた例である。表示方法の切り替えは、TimeSlice を奥行き方向にスライドさせた状態 (Fig. 3 (a)) で、ユーザ視点を時間軸方向から TimeSlice を見るように移動し (Fig. 3 (b)), 正射影モードで描画することで実現する (Fig. 3 (c)).

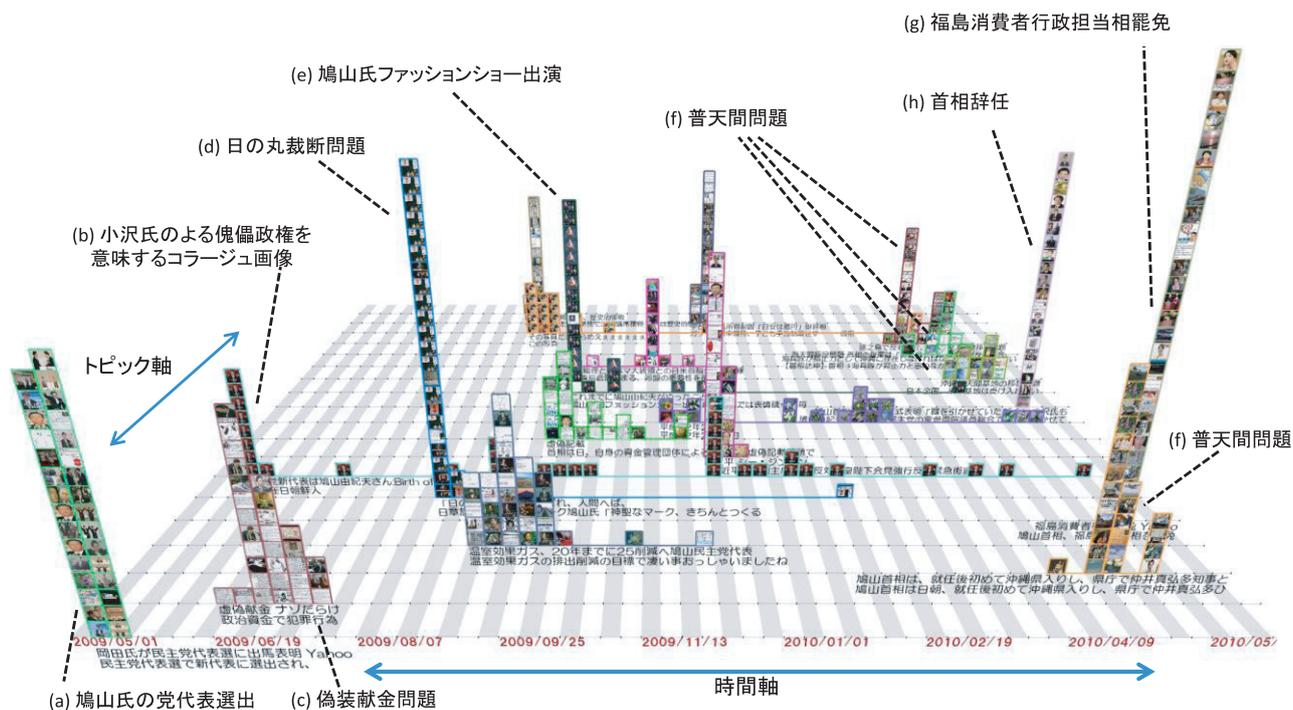


Fig. 4 ブログ画像のクラスタリングを用いたトレンド可視化：政治・社会問題に関する話題変遷を可視化事例（「鳩山由紀夫」を検索語に用いた場合）.

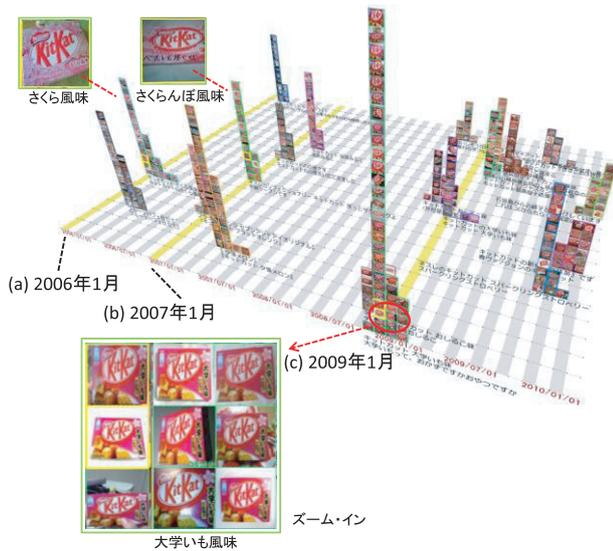


Fig. 5 ブログ画像のクラスタリングを用いたトレンド可視化：話題商品の推移可視化事例（キットカットの例）。

4. ブログ画像のクラスタリングを用いたトレンド可視化システム

ソーシャルメディア上では、テキストの代用として積極的に画像を用いることで、文章だけでは伝えきれない、その時々話題および興味を視覚的に発信することも多く、社会分析において、これらの画像情報を追跡した分析は不可欠となる。この章では、流行画像を通してソーシャルメディア上のトレンドを探索するための可視化システム⁵⁾を紹介する。

このシステムでは、まず、ソーシャルメディア上の話題画像の遷移を把握可能にするための情報抽出手法として、画像特徴、テキスト特徴、および時間情報を用いたクラスタリングを行い画像クラスタを抽出している。抽出された画像クラスタをそれぞれトピックとして扱い、画像ヒストグラムとして時間軸上に可視化する。このシステムを用いることで、政治家、政治問題などに関する社会活動イベント、季節ごとにラインナップが変化する商品の人気推移、あるいは、企業が行うキャンペーンの

話題変遷などが視覚的に把握可能になる。

Fig. 4 では、検索語として「鳩山由紀夫」を用いた場合の上位 20 クラスタを手前から順に可視化している。画像は 1 週間ごとに集約し、画像ヒストグラムとして時系列上に表示している。図の例では、(a) 鳩山氏の党代表選出から、(c) 偽装献金問題、(d) 日の丸裁断問題などの政治的な話題がクラスタリングされ、順に出現している。さらに、(f) 普天間問題、(g) 福島消費者行政担当相罷免を経て (h) 首相辞任といった首相辞任に直接関係した話題がほぼ同時期に出現していることがわかる。また、(d) (e) のようなあまり報道に使用されないようなイベントや集会、あるいは (b) のようなソーシャルメディアのみで言及される話題を発見しその内容を視覚的に把握できていることがわかる。特に (b) の小沢氏による傀儡政権を意味するコラージュ画像は鳩山氏が党代表に選出された直後から首相を辞任するまで長期にわたり政権に何かが起こるたびに繰り返し出現しており、ウェブ上での政権に対するイメージが表現されている。

Fig. 5 の例では、ブログにおける話題推移の一例として「キットカット」を検索語とした商品トレンドの可視化例を示す。クラスタ数は上位 20 クラスタ、画像は 1 ヶ月ごとに集約している。図の例では、「キットカット」に関するフレーバの流行具合が確認出来る。ここから (a) - (c) のようにほぼ毎年 1 月ごろに集中して様々なフレーバが流行していることが見て取れ、特に「サクラ味」「さくらんぼ味」「大学芋味」などが大きく流行している。元ブログの内容を確認することで、「きっと勝つ」「桜咲く」「大学」「金星（大学芋の色）」という語呂合わせ的な意味合いから大学受験シーズンに受験生応援商品として支持を集めていたと確認できている。

5. 放送・Web 間話題伝搬追跡システム

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、どのようにその話題が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。放送メ

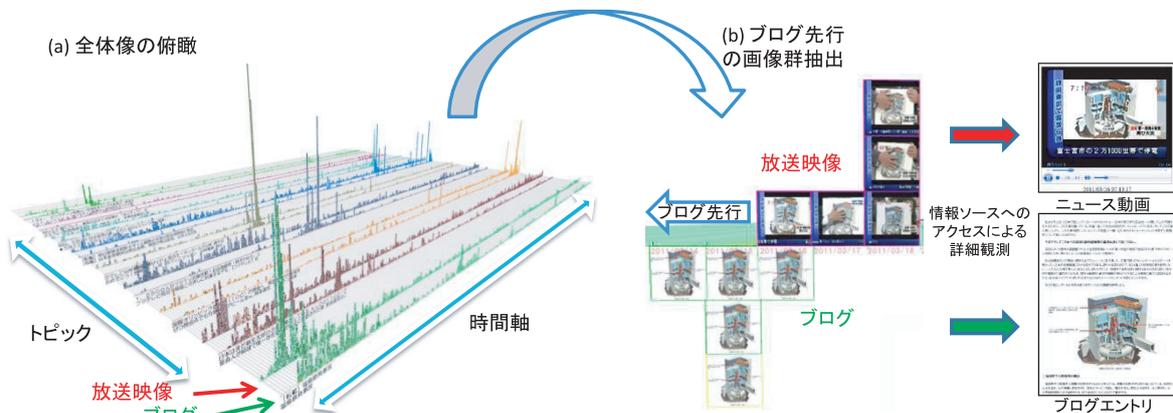


Fig. 6 画像クラスタ照合を用いた放送・Web 間話題伝搬追跡システム。

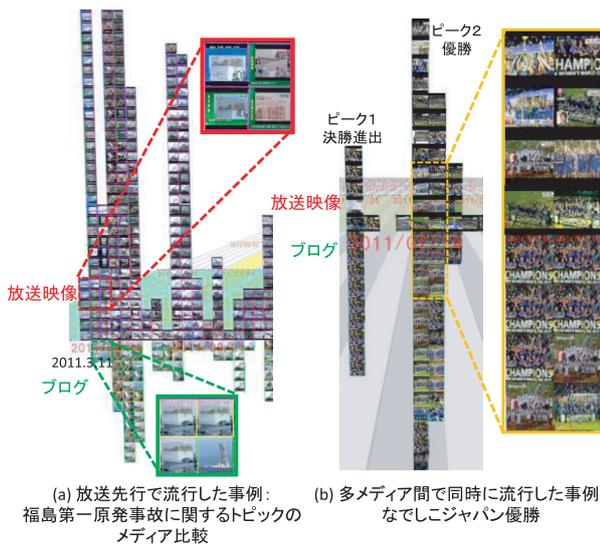


Fig. 7 放送先行で流行した事例と同時に流行した事例。

ディアおよびウェブでは、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、複数メディア画像を相補的に用いた可視化・探索環境が必須となる。著者らは、ウェブおよび放送映像から抽出された時系列画像群頻度変化を並列可視化することで、メディア間の影響分析を視覚的に可能にする放送・Web間の話題伝搬追跡システムを構築した⁶⁾。

Fig. 6の例では、震災、原発、オリンピック等のトピックキーワードでブロッガーアーカイブから抽出した画像群を画像類似度でクラスタリングし、さらに放送映像アーカイブから各ブログ画像クラスタに類似する画像群を抽出し、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 (Fig. 6 (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、コサイン類似度などから様々な反応差をもつ画像クラスタを対話的に抽出し詳細を探索することが可能である。同図の(b)の例では、ブログ先行の画像群を抽出している。ここで抽出された画像クラスタの画像は、MITの研究者による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。この画像が、まずブログ上でMIT研究者による解説が話題なることで広まり、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様子が見て取れる。

Fig. 7の(a)は、放送映像で先行して出現した画像群を抽出して時間軸上に可視化した一例である。ここから、原発の空撮映像が報道番組中でまず繰り返し使用され、その後、ブログ上で広まったことが見て取れる。Fig. 7の(b)は、放送メディアとブログで同時に流行した画像クラスタを抽出した例である。この例では、「なでしこジャパン」の女子ワールドカップ2011における優勝に関連した画像クラスタ群が抽出されおり、流行のピークが2度現れていることが分かる。1度目のピークは準決

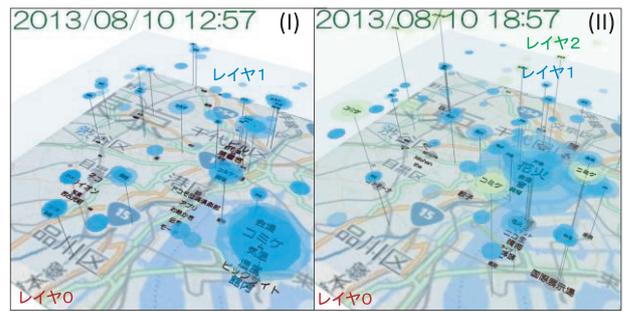


Fig. 8 複数の大規模集客イベント（コミックマーケットと東京湾大華火祭）発生時の可視化事例。

勝に勝ち決勝進出を決めた日であり、2度目のピークは、優勝した日とそれに続く数日間である。どちらの場合も放送映像クラスタとブログ画像クラスタ両方で同時にピークが現れており、メディアの種類によらず全体としてポジティブな盛り上がりを見せていたことが伺える。

6. 位置参照表現に着目した実世界イベントの時空間可視化システム

東京などの大都市では、様々な種類のイベント（大規模集客イベント、交通事故、地震・台風など）が日々発生しており、それらの時間・空間的影響範囲および注目度などを知ることは都市計画などの観点から重要になってくる。著者らは、大都市で日々発生するイベントの影響、時空間的な広がりを理解するための、マイクロブログストリーム中の位置参照表現に着目した実世界イベント時空間可視化システムを構築した⁷⁾。

このシステムでは、まず、ツイッターストリーム中の各投稿中に現れる地名や施設名などの位置参照表現を認識する。次に、投稿中の各単語を、認識した地名や施設名の位置に紐付けて、局所的に観測された単語をローカルイベントとして認識する。さらに、限定された場所で起こるローカルイベント群から広範囲に広がるグローバルイベント群を区別し抽出し、それらを3次元空間の別レイヤにワードクラウドとして可視化する。この3次元空間における多層地理空間ワードクラウド表現により個々のイベントの探索、イベント間の関係の探索、およびイベントが時間的・空間的に変化していく様子が視覚的に理解可能になる。

Fig. 8は、二つのグローバルイベントが発生した日の例である。一つ目のイベントは東京ビッグサイトで行われるコミックマーケット（コミケ）で、コミケの開催日には、さまざまな場所で連動イベントが行われたり、コミケの行き帰りの人々およびそれを目撃した人々が様々な場所でコミケに関連した投稿を行う。そのため、朝の早くからグローバルイベントとして抽出され、レイヤ1上に表示されている (Fig. 8 (I))。二つ目のイベントは晴海埠頭で行われた東京湾大華火祭で、夕方過ぎからグローバルイベントとして抽出されている。そのため、Fig. 8 (II)では二つのレイヤ上にそれぞれのグローバル

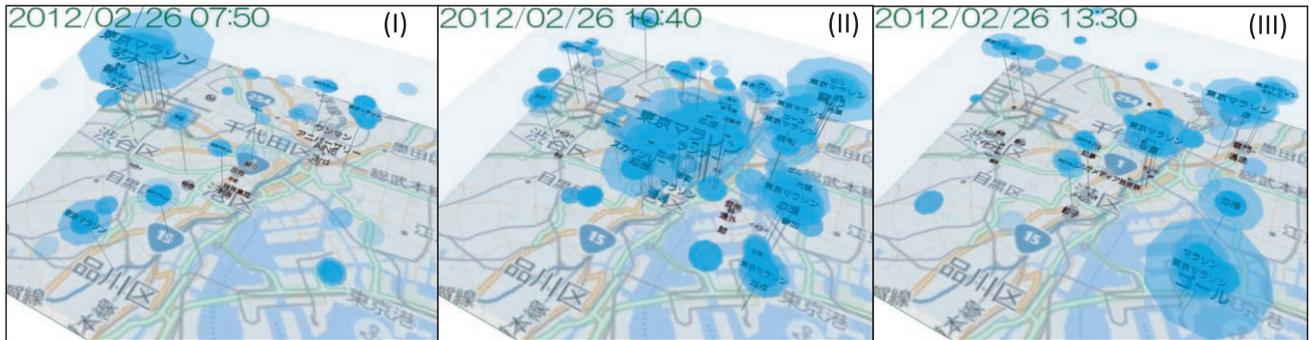


Fig. 9 東京マラソン開催日におけるイベント変遷の可視化.

イベントが可視化されていることが分かる。また、地表面であるレイヤ0上には、ローカルイベントである"ビル", "落書き", "取り壊し"などが表示されていることが確認できる。

Fig. 9 は東京マラソン 2012 が開催された日におけるイベント変遷を可視化した例である。東京マラソンでは約 35,000 のランナーが参加しており 100 万人以上の人々が沿道から観戦をしている。東京マラソンのコースは東京の西から東への移動、さらに南北間での 2 度の折り返しを含み、完走する多くのランナーは約 3 時間から 6 時間でゴールする。また、多くの人々がランナーを応援するために主要観戦ポイント間を移動し、状況をツイッターなどに投稿している。Fig. 9 (I) から (III) では、"東京マラソン"という大きなグローバルイベントとそれに関連する様々なイベントとそれらの時間変遷を観測できる。それぞれの時間のスクリーンショットにおけるイベントのサイズと場所は、その時間におけるランナーの走行位置を反映している。関連イベントのサイズと場所からイベントの詳細状況を知ることができる。

7. まとめ

本稿では、ソーシャルメディアから社会の変化を読み解く助けとなるような可視化・探索システム群を紹介した。詳細とさらなる事例については既発表論文^{3), 4), 5), 6), 7)}を参照されたい。GPS などの物理センサに対して、ソーシャルメディアをソーシャルセンサとみなし、様々な種類の大規模センサデータを統合し、実世界を制御しうる有意義な情報を探索可能にしようとする動きも活発になっている。本稿では割愛したが、著者らも、公共交通網のセンサデータとツイッターデータを相補的に利用することによる事故、災害などの影響分析を行うシステムを構築している⁸⁾。今後、これらの仕組み

が、マーケティング、社会分析、都市計画などさまざまな目的に実際に利用されていくことを期待している。

参考文献

- 1) 伊藤正彦, 豊田正史, 喜連川優: 構造を持つ時系列情報に關する 3 次元可視化フレームワーク, 第 19 回インタラクティブシステムとソフトウェアに関するワークショップ, (2011) pp. 42-47.
- 2) 伊藤正彦, 豊田正史, 喜連川優: 多メディア間の話題探索のための時系列画像 3 次元可視化システム, 情報処理学会論文誌: データベース, Vol.8, No.1, (TOD65), (2015) pp. 27-44.
- 3) M. Itoh, M. Toyoda, M. Kitsuregawa: An Interactive Framework for Visualizing Time-series of Web Graphs in a 3D Environment, In Proceedings of the 14th International Conference on Information Visualization, (2010) pp. 54-60.
- 4) M. Itoh, N. Yoshinaga, M. Toyoda, M. Kitsuregawa: Analysis and Visualization of Temporal Changes in Bloggers' Activities and Interests, In Proceedings of 5th IEEE Pacific Visualization, (2012) pp. 57-64.
- 5) M. Itoh, M. Toyoda, M. Kitsuregawa: Visualizing Time-Varying Topics via Images and Texts for Inter-Media Analysis, In Proceedings of the 17th International Conference on Information Visualization, (2013) pp. 568-576.
- 6) M. Itoh, M. Toyoda, C. Zhu, S. Satoh, M. Kitsuregawa: Image Flows Visualization for Inter-Media Comparison, In Proceedings of 7th IEEE Pacific Visualization, (2014) pp. 129-136.
- 7) M. Itoh, N. Yoshinaga, M. Toyoda: Spatio-temporal Event Visualization from a Geo-parsed Microblog Stream, In Proceedings of the 21st ACM International Conference on Intelligent User Interfaces, (2016) (poster).
- 8) M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, M. Kitsuregawa: Visual Fusion of Mega-City Big Data: An Application to Traffic and Tweets Data Analysis of Metro Passengers, In Proceedings of 2014 IEEE International Conference on Big Data, (2014) pp. 431-440.