

テレビアーカイブを用いたアナリティクスのための 関連ニュースショット検出

片山 紀生^{1,2,a)} 孟 洋^{1,2,b)} 佐藤 真一^{1,c)}

概要: テレビアーカイブの解析は、防災・安全など社会的現象を調査・検証する手段として有用であるが、その際には、対話的な解析プロセスの中で、キーワード検索などによって得られたショットを基点として、関連するショットを探索することが求められる。本稿では、言語、音声、映像を組み合わせたマルチモーダルな探索を行うことにより、関連ショットを効果的に検出できることを示す。

キーワード: テレビアーカイブ, データアナリティクス, 防災, 安全, ニュースショット検索

News Shot Retrieval for Data Analytics with Television Archive

Abstract: Data analytics of news contents in television archive is useful for disaster and risk prevention. During the interactive process of analyzing television archive, it is required to retrieve relevant news shots for a news shot that is found by some search or visualization method. This paper shows that relevant shots can be retrieved using multi-modal search that includes textual, audio, and visual search.

Keywords: Television Archive, Data Analytics, Disaster Prevention, Risk Prevention, News Shot Retrieval

1. はじめに

テレビアーカイブの解析は、防災・安全など社会的現象を調査・検証する手段として有用であるが、その際には、対話的な解析プロセスの中で、キーワード検索などによって得られたショットを基点として、関連するショットを探索することが求められる。本稿では、言語、音声、映像を組み合わせたマルチモーダルな探索を行うことにより、関連ショットを効果的に検出できることを示す。

2. 背景

2.1 テレビアーカイブの活用

近年の計算機技術の進歩により、大量の映像をデジタル記憶することが可能になり、テレビ放送を長期間に渡って蓄積したテレビアーカイブが作られるようになっていく。

立情報学研究所でも、学術研究目的のテレビアーカイブである NII TV-RECS [1], [2] を構築しており、2009年8月より東京地区の地上波7チャンネルの放送映像を連日24時間蓄積している。また、映像情報に加えてテキスト形式の情報として、字幕放送の文字字幕(クローズドキャプション)および番組情報も蓄積している。このようなテレビアーカイブを構築する動きは海外でも進められており、アメリカ合衆国の Vanderbilt Television News Archive, TV News Archive (Internet Archive), UCLA Library NewsScape, フランスの INA MEDIAPRO Television Collection, オランダの Beeld en Geluid (Sound and Vision) などがある。このようなテレビアーカイブには、様々な災害、事件、事故に関するニュースが蓄積されており、防災・安全意識を高めるためのリソースとして活用することが考えられる。

インターネットの普及により若者のテレビ離れが進んでいることが指摘されており、情報獲得手段としてのテレビ放送の役割は、相対的に小さくなっていると考えられる。しかし、テレビ放送は、同報性や広域性の点では現在でもメリットがあり視聴者も多い。視聴者が多いことで、衆人環視による評価の目も厳しくなるため、品質や信頼が保た

¹ 国立情報学研究所
National Institute of Informatics

² 総合研究大学院大学
The Graduate University for Advanced Studies, SOKENDAI

a) katayama@nii.ac.jp

b) mo@nii.ac.jp

c) satoh@nii.ac.jp

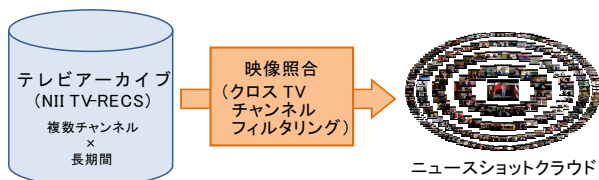


図 1 ニュースショットクラウド
 Fig. 1 News Shot Cloud.

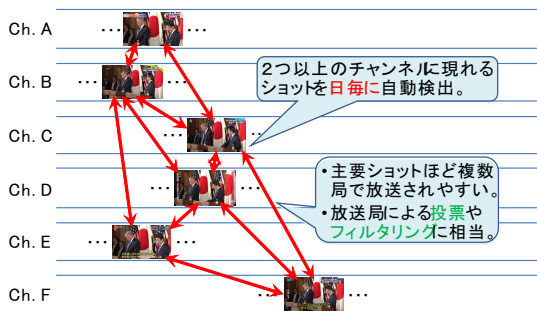


図 2 クロス TV チャンネルフィルタリング
 Fig. 2 Cross TV-Channel Filtering.

れやすい面もある。実際、総務省の調査によると、テレビに対する信頼度は、新聞に次いで高いことが示されている [3]。また、日本新聞協会による調査においても、テレビは新聞に次いで信頼度の高いメディアとなっており、かつ、コロナ禍で接触頻度が最も増えたメディアとされている [4]。そのため、特に災害時等での情報伝達では、現在でも大きな役割を担っていると考えられる。

2.2 マルチメディア・データ・アナリティクス

今世紀に入り、インターネット、および、センサーや記憶装置などの電子デバイスの高性能化により、大量のデータを収集・蓄積できるようになり、いわゆるビッグデータの時代が到来している。そして、それに呼応して、2010 年頃より、ビッグデータを解析するための基盤として、Data Analytics や Data Science に注目が集まるようになっていく [5]。また、マルチメディアデータを対象としたアナリティクスとして、Multimedia Data Analytics, あるいは、Multimedia Analytics にも関心が高まり、大量のマルチメディアデータを解析および可視化するための基盤として、方法論やプラットフォームの研究開発が進められている [6], [7], [8], [9], [10], [11], [12], [13]。

一方、テキストデータを対象としたビッグデータの解析では、早くから、地震や台風などの災害発生時の SNS データ (ツイートなど) を用いた技術や応用が注目されており、人々の振舞いを把握・解析する上で、位置情報付きのツイートなどが有効であることが示されている [14], [15], [16], [17]。

そのような中、国立情報学研究所では、学術研究目的のテレビアーカイブである NII TV-RECS [1], [2] を用いたマルチメディアアナリティクスの研究に取り組んでおり、

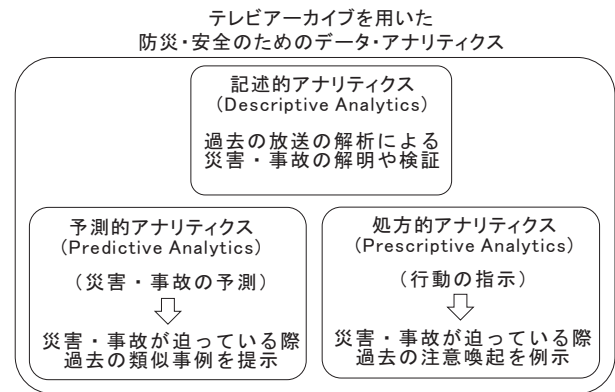


図 3 テレビアーカイブを用いたデータアナリティクス
 Fig. 3 Data analytics with TV archive.

2011 年の東日本大震災の際には、「NII 研究用テレビジョン放送アーカイブを用いた東日本大震災の社会的影響の学術的分析」というテーマで共同研究を公募し、震災時のテレビ放送について、幅広い視点での解析が行われた [18], [19]。筆者らも、映像の構図による分類に基づく震災テレビ報道の分析を行った [20]。また、2018 年には西日本豪雨 (平成 30 年 7 月豪雨) の際のテレビ報道の傾向解析 [21] を行うとともに、2019 年にはテレビアーカイブを防災・安全意識を維持・高めるための道具として応用する可能性について考察した [22]。

2.3 ニュースショットクラウド

我々はこれまでに、テレビアーカイブを俯瞰的に閲覧するための手法として、「ニュースショットクラウド」と名付けた手法について研究を進めてきた [23], [24]。ニュースショットクラウドは、主要なニュースショットを映像照合によって検出し、Web で用いられているタグクラウドのように、雲状に配置したものである (図 1)。映像照合の目的は、複数のチャンネルで共通に放送されたニュースショットを発見することにある。この手法では、「視覚的に有用なショットほど複数のチャンネルで共通して放送されやすい」という仮説に基づき、ニュース番組等で放送されているニュースショットの中で、同一またはほぼ同一のものがいくつのチャンネルで放送されているかを映像同士の照合によって検出する (図 2)。そして、出現チャンネル数が多いニュースショットほど、より視覚的に有用なショットであると見なして検出する。この処理のことを「クロス TV チャンネルフィルタリング」と呼んでおり、複数チャンネルを活用した情報フィルタリングと見ることができ、放送局によるソーシャルフィルタリング、または、放送局による人気投票とも見なせるものである。

3. テレビアーカイブを用いたアナリティクス

3.1 防災・安全のためのアナリティクス

データアナリティクスは、しばしば以下の 3 つの段階に

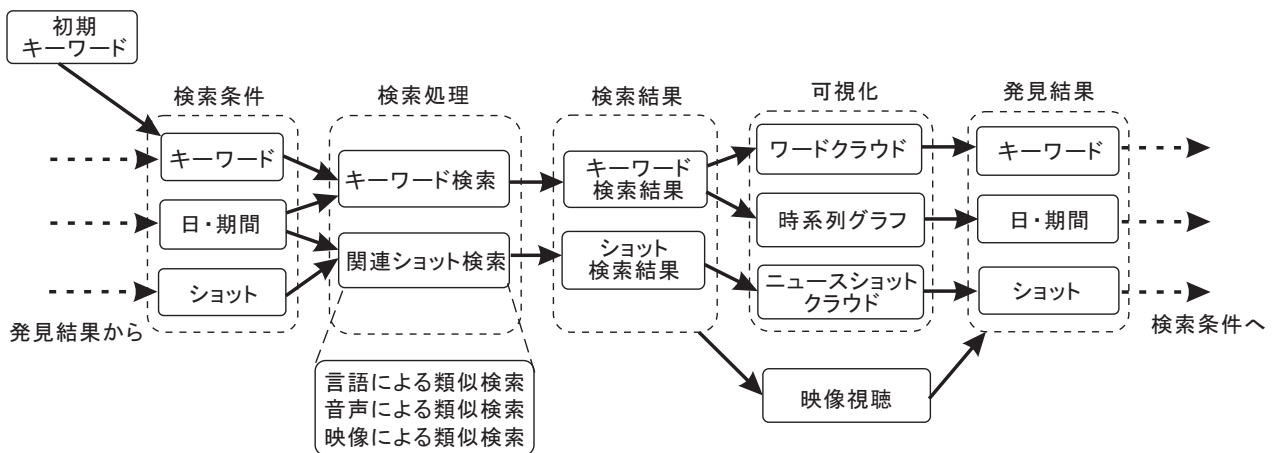


図 4 対話的プロセスの流れ

Fig. 4 Flow of interactive process.

分けて考えられる [12], [25].

- 記述的アナリティクス (Descriptive Analytics)
- 予測的アナリティクス (Predictive Analytics)
- 処方的アナリティクス (Prescriptive Analytics)

記述的アナリティクスは、過去に何が起きたかを解析するものであり、予測的アナリティクスは、将来に何が起きるかを予測するものであり、処方的アナリティクスは、何をすべきかを提案するものである。これを、テレビアーカイブを解析対象とし、かつ、防災・安全を目的としてどのようなことが可能か考えると以下のようなになる。

まず、記述的アナリティクスは既に広く行われていることであり、大きな災害や事故が起きた場合にテレビアーカイブを解析することで何が起きたかを明らかにできる。実際、前節で述べたとおり、東日本大震災や西日本豪雨を対象とした解析が行われている。

次に、予測的アナリティクスは将来を予測することになるが、災害や事故をテレビアーカイブを用いて正確に予測することは用意ではない。しかし、何か災害や事故が迫っている際に、過去の類似事例を探索することで、起こるかもしれないことを見つけ出すことは可能だと考えられる。

そして、処方的アナリティクスは、データアナリティクスの3つの段階の中で、最も進んだ段階、すなわち、最も難しい段階と言われており、テレビアーカイブを用いるだけではできることが限られていると考えられる。しかし、予測的アナリティクスによって過去の類似事例が見つかった場合には、その際にどのような注意喚起が行われていたかを抽出すれば、どのような行動を取るべきかについて参考となる情報を提示できる可能性がある。

これらのことから、テレビアーカイブを用いた防災・安全のためのアナリティクスとしては、図3のように取り組むことが考えられる。予測的アナリティクス、および、処方的アナリティクスについては、完全な予測や完全な処方達成できないと考えられるが、参考となる情報を提示す

ることは可能であると考えられる。

3.2 認知バイアスの補正と信頼性の検証

上記の記述的／予測的／処方的アナリティクスは、データアナリティクスの効果として代表的なものであり、災害や事故の解明や予防に役立つことが期待できる。一方、テレビアーカイブを用いたデータアナリティクスの場合、これらと異なる観点での効果として、以下の2点がある。

- 認知バイアスの補正
- 信頼性の検証

近年、情報獲得手段の多様化や社会の分断化・高齢化などにより、確認バイアスや正常性バイアスなどの認知バイアスが社会問題になっている。そのような認知バイアスを緩和するには、当事者の意識改革も必要である一方、多様な情報に接することが重要である。テレビアーカイブを用いたデータアナリティクスを活用することで、テレビを短時間視聴するだけでは得られない多様な情報を提供できる可能性があり、認知バイアスを補正するための手段のひとつとしての有効性を期待できる。

他方、テレビに対しては不信感もはびこっており、2020年の米国の大統領選挙では、米国の放送局の偏向やニュースの信憑性が問題になった。また、日本国内でも、特に新型コロナに関する報道について、マスコミが煽っているという意見や都合の悪い情報を隠している(伏せている)という意見が出るなどメディアの信頼性を疑問視する意見が出ている。2020年11～12月の朝日新聞による世論調査 [26] では、テレビ、新聞を信頼している人は、メディアが必要な情報を提供していると7割の人が回答しているが、SNSを信頼している人は、41%の人が提供していると回答し、56%の人が提供していないと回答したことが報じられている。このような不信は、漠然とした疑念や不安から生じている場合も少なくなく、テレビアーカイブのアナリティクスを用いることでその妥当性を検証できると考えられる。

例えば、よくある例として、「テレビでは〇〇について意図的に取り上げていないのではないか？」というものがあるが、そのような疑念が正当かどうかは、テレビアーカイブを用いて容易に確かめることが可能である。

4. 関連ニュースショット検出

4.1 対話的プロセス

アナリティクスによる情報の探索は、視覚化を行いながらの対話的なプロセスになる。テレビアーカイブを対象に、ワードクラウドやニュースショットクラウドを用いながら探索する過程は、図4のようになる。検索条件として、キーワード、ショット、期間を使いながら検索を行うとともに、検索結果を可視化したり、映像を視聴することによって、新たなキーワードやショットを発見しながら探索を進める。テレビアーカイブは一般に長期間に渡ってテレビ放送を蓄積しているため、対象とする期間を適切に絞り込むことも重要である。

4.2 マルチモーダルなショット検索

ショット検索については、言語、音声、映像を組み合わせたマルチモーダルな検索が有効である。というのも、ショット間の関連性は、言語、音声、映像のすべてが類似しているとは限らず、一部のみしか類似していないことも多いからである。例えば、同じ時刻、同じ場所の映像であっても、撮影方向が異なれば映像は異なったものになる。また、逆に、言語情報や音声情報が異なっても、映像のみの類似性からショット間の関連性が見つかることもある。

また、類似性を評価する際には、一致の条件を使い分けることも重要である。検索したい対象が十分に絞り込まれている段階では、一致の条件を厳しくし、ほぼ一致するもののみを検索することが考えられるが、逆に、検索したい対象が絞り込まれておらず幅広く探索したい段階では、一致の条件を緩くすることが考えられる。

このように関連ショットの検索においては、多角的な検索が可能であり、検索の目的や段階に応じて適応的に取捨選択することが必要である。

5. 解析事例

5.1 コロナ禍でのテレビ報道

コロナ禍のもと2020年1月以降、テレビでは新型コロナについて多様な報道が行われている。図5は、ニュース番組および情報番組の字幕放送の中での「新型コロナ」という語の出現数、および、新規陽性者数をプロットしたものである。この図からわかるとおり、新型コロナに関する報道は第一波となった2020年4月が最も多くなっているが、それ以降も持続的に報道が続いていることがわかる。図6は、この期間の字幕放送の中で「新型コロナ」の周辺

に頻出した語をワードクラウドによって視覚化したものである。新型コロナに関連して、様々なトピックが取り上げられてきたことがわかる。

5.2 不顕性感染への警戒

新型コロナについては、不顕性感染、すなわち、感染者が無症状のことがあり、しかも、その状態で周囲に感染させることが大きな特徴となっており、そのことがこのウィルスの感染予防を難しくしている。

そこで、ニュース番組および情報番組での「無症状」の出現数をプロットすると図7のようになり、第一波より数ヶ月前の2020年1月末から現れていることがわかる。最初のピークは2020年1月30日に現れており、この日の報道内容をニュースショットクラウドによって確認すると、図8のようになっており、厚生労働省での記者会見が最も主要なショットとなっていた。この記者会見では、帰国者の中に無症状の感染者が見つかったことが報告されており、そのような人が周囲に感染させるかどうか話題になっていた。このトピックについて関連ショットの検出を行うと、図9のショットが得られた。NHK、日本テレビ、TBSは同じ発言を異なる角度から撮影したものであり、テレビ朝日は発言したタイミングは異なるがほぼ同じ内容のものだった。さらに、NHKでは、厚生労働省の担当者と専門家とのやりとりのショットも含まれており、当時は無症状者が感染させるかどうかについて厚生労働省が慎重な立場を取っていたため、専門家との間で意見が食い違っている様子が報じられていた。

このように、日本国内では第一波の数ヶ月前から無症状感染について話題になっていたことがわかる。このことは、ツイート数でも確かめることができ、「無症状でも」を含むツイートの件数をプロットすると図10のようになっており、テレビ報道と同様に2020年1月末から話題になっていることがわかる。無症状でも感染させることは新型コロナの大きな特徴であり、そのような特徴が早い段階で把握されていたことは興味深い結果である。また、そのような事象を明確に検証できることは、テレビアーカイブを用いたデータアナリティクスの有用性を示すものである。

6. まとめと今後の課題

本稿では、テレビアーカイブを用いたデータ・アナリティクスについて、どのような形で対話的プロセスが構築されるかを示すとともに、その中での関連ショット検出の役割と効果について考察した。そして、解析事例としてコロナ禍でのテレビ報道に着目し、不顕性感染を例に関連ニュースショット検出の有効性を検証した。テレビアーカイブを用いた防災・安全のためのデータアナリティクスは、災害・事故の解明・予防にとって有用であると期待できる。今後は、解析プロセスのさらなる支援を目的として、関連

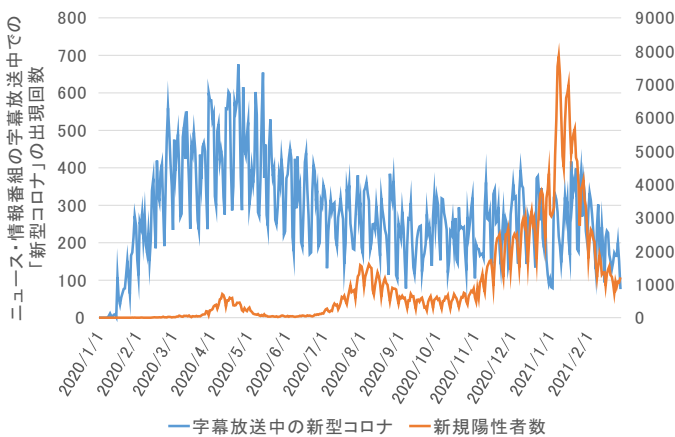


図 5 字幕放送中の「新型コロナ」の出現数

Fig. 5 Occurrence of “COVID” in closed captions.

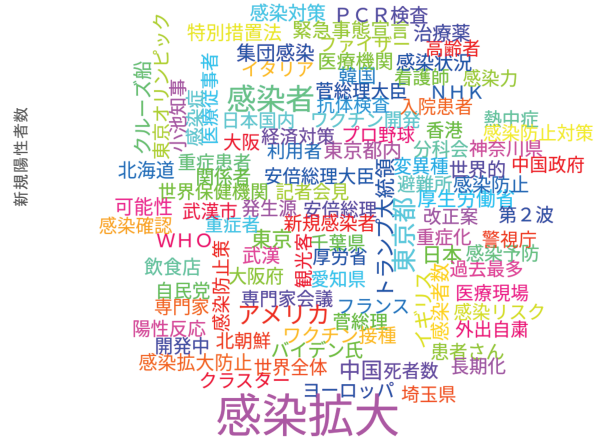


図 6 「新型コロナ」の周辺での頻出語のワードクラウド

Fig. 6 Frequent words near “COVID”.

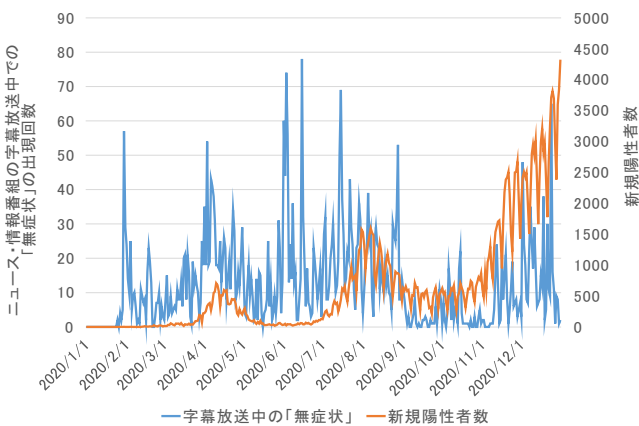


図 7 字幕放送中の「無症状」の出現数

Fig. 7 Occurrence of “without symptoms” in closed captions.



図 8 2020年1月30日のニュースショットクラウド

Fig. 8 News shot cloud of 2020/01/30.

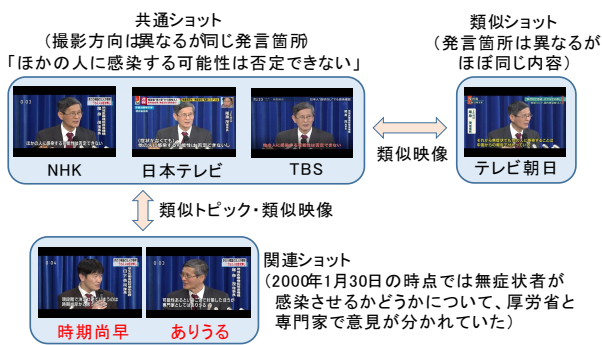


図 9 無症状者でも感染するかについての関連ショット

Fig. 9 Relevant shots on the patients without symptoms.

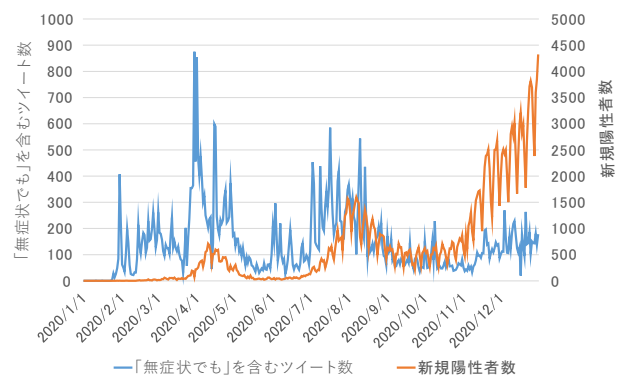


図 10 「無症状でも」を含むツイート数

Fig. 10 Tweets containing “without symptoms”.

ショット検出の精度向上や視覚化の高機能化を進める考えである。

謝辞 本研究は、JST CREST（ Grant 番号 JP-MJCR19F4）および JSPS 科研費（JP18K11386）の支援を受けました。

参考文献

- [1] 片山紀生, 孟 洋, 佐藤真一: 映像インデクシング研究のための大規模放送映像アーカイブシステムの試作, 情報処理学会研究報告 (DBS), 2002(41), pp. 17–23 (2002).
- [2] Katayama, N., Mo, H., Ide, I. and Satoh, S.: Mining Large-Scale Broadcast Video Archives towards Inter-Video Structuring, *Pacific Rim Conference on Multimedia (PCM2004) LNCS, vol.3332*, pp. 489–496 (2004).
- [3] 総務省情報通信政策研究所: 情報通信メディアの利用時間と情報行動に関する調査報告書 (2020).

- [4] 日本新聞協会: 新型コロナウイルスとメディア接触・信頼度調査 (2020).
- [5] Cao, L.: Data Science: A Comprehensive Overview, *ACM Comput. Surv.*, Vol. 50, No. 3, pp. 43:1–43:42 (2017).
- [6] Chinchor, N. A., Thomas, J. J., Wong, P. C., Christel, M. G. and Ribarsky, W.: Multimedia Analysis + Visual Analytics = Multimedia Analytics, *IEEE Computer Graphics and Applications*, Vol. 30, No. 5, pp. 52–60 (2010).
- [7] Jónsson, B., Worring, M., Zahálka, J., Rudinac, S. and Amsaleg, L.: Ten Research Questions for Scalable Multimedia Analytics, *MultiMedia Modeling (MMM 2016)*, *LNCS*, vol.9517, pp. 290–302 (2016).
- [8] Gravier, G., Ragot, M., Amsaleg, L., Bois, R., Jadi, G., Jamet, E., Monceaux, L. and Sébillot, P.: Shaping-Up Multimedia Analytics: Needs and Expectations of Media Professionals, *MultiMedia Modeling (MMM 2016)*, *LNCS*, vol.9517, pp. 303–314 (2016).
- [9] Wu, Y., Cao, N., Gotz, D., Tan, Y.-P. and Keim, D. A.: A Survey on Visual Analytics of Social Media Data, *IEEE Trans. on Multimedia*, Vol. 18, No. 11, pp. 2135–2148 (2016).
- [10] Kurzhals, K., John, M., Heimerl, F., Kuznecov, P. and Weiskopf, D.: Visual Movie Analytics, *IEEE Trans. on Multimedia*, Vol. 18, No. 11, pp. 2149–2160 (2016).
- [11] Renoust, B., Le, D.-D. and Satoh, S.: Visual Analytics of Political Networks From Face-Tracking of News Video, *IEEE Trans. on Multimedia*, Vol. 18, No. 11, pp. 2184–2195 (2016).
- [12] Pouyanfar, S., Yang, Y., Chen, S.-C., Shyu, M.-L. and Iyengar, S. S.: Multimedia Big Data Analytics: A Survey, *ACM Comput. Surv.*, Vol. 51, No. 1, pp. 10:1–10:34 (2018).
- [13] Zhang, W., Yao, T. and Zhu, S.: Deep Learning-Based Multimedia Analytics: A Review, *ACM Trans. on Multimedia Computing, Communications, and Applications* (2019).
- [14] Imran, M., Castillo, C., Diaz, F. and Vieweg, S.: Processing Social Media Messages in Mass Emergency: A Survey, *ACM Comput. Surv.*, Vol. 47, No. 4, pp. 67:1–67:38 (2015).
- [15] Li, T., Xie, N., Zeng, C., Zhou, W., Zheng, L., Jiang, Y., Yang, Y., Ha, H.-Y., Xue, W., Huang, Y., Chen, S.-C., Navlakha, J. and Iyengar, S.: Data-Driven Techniques in Disaster Information Management, *ACM Comput. Surv.*, Vol. 50, No. 1, pp. 1:1–1:45 (2015).
- [16] Alfarrarjeh, A., Agrawal, S., Kim, S. H. and Shahabi, C.: Geo-Spatial Multimedia Sentiment Analysis in Disasters, *IEEE Intl. Conf. on Data Science and Advanced Analytics (DSAA 2017)*, pp. 193–202 (2017).
- [17] Nazer, T. H., Xue, G., Ji, Y. and Liu, H.: Intelligent Disaster Response via Social Media Analysis A Survey, *ACM SIGKDD Explorations Newsletter*, Vol. 19, No. 1, pp. 46–59 (2017).
- [18] 高野明彦, 吉見俊哉, 三浦伸也: 311 情報学 —メディアは何をどう伝えたか, 岩波書店 (2012).
- [19] 伊藤 守: テレビは原発事故をどう伝えたのか, 平凡社 (2012).
- [20] 片山紀生, 孟 洋, 佐藤真一: 視覚的情報の役割に着目したニュースショット分類による震災テレビ報道の分析, 情報処理学会研究報告, 2012-IFAT-106(6), pp. 1–8 (2012).
- [21] 片山紀生, 孟 洋, 佐藤真一: マルチメディアアナリティクスによる防災・災害テレビ報道の傾向解析, 電子情報通信学会技術研究報告 (PRMU) 118(513), No. 109–112 (2019).
- [22] 片山紀生, 孟 洋, 佐藤真一: 防災・安全を目的とする記憶補完支援へのテレビアーカイブの応用可能性, 電子情報通信学会技術研究報告 (PRMU) 119(481), pp. 187–189 (2020).
- [23] Katayama, N., Mo, H. and Satoh, S.: News Shot Cloud: Ranking TV News Shots by Cross TV-Channel Filtering for Efficient Browsing of Large-Scale News Video Archives, *Advances in Multimedia Modeling (LNCS vol. 6523/2011)*, pp. 284–295 (2011).
- [24] Katayama, N., Mo, H. and Satoh, S.: Enhanced Visualization of News Shot Cloud with Employing Circular Layout, *International Symposium on Visual Information Communication and Interaction (VINCI 2015)*, pp. 166–167 (2015).
- [25] Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. and Buyya, R.: Big Data computing and clouds: Trends and future directions, *Journal of Parallel and Distributed Computing*, Vol. 79-80, pp. 3–15 (2015).
- [26] 朝日新聞: コロナ情報、信じていいか「不安」74% 朝日世論調査 (2021.1.8).