

放送素材の利活用のための可視化手法

Visualization of Broadcast Materials for effective utilization of TV Program Production

宮崎 太郎[†] 石渡 太智[†] 遠藤 伶[†] 望月 貴裕[†] 後藤 淳[†] 霜山 文雄[‡] 内田 明香[‡] 田中 清隆[‡]
 Taro Miyazaki Taichi Ishiwatari Rei Endo Takahiro Mochizuki
 Jun Goto Fumio Shimoyama Sayaka Uchida Kiyotaka Tanaka

1. はじめに

放送局は、過去の番組制作のために取材してきた膨大な映像やテキストなどのリソースを保有している。これらは貴重な一次資料であり、取材から長い年月が経過してからも頻りに利用される場合も多い。一方で、その量が膨大であることから、リソースそのものを俯瞰して分析したりするのは難しかった。これらのリソースを分析し、活用することで、番組制作の効率化や、従来とは異なる視点からの番組制作に結び付けられる可能性があることから、分析技術への要望が高くなっている。放送素材の一つであるニュース記事の可視化手法は、自己組織化マップを用いた可視化手法[1]や、lanes metaphor による可視化手法[2]など多くの研究がなされている。しかし、多くの場合は番組制作者が使用することは想定していなかった。

そこで我々は、NHK 知財センターで保有している過去のニュース素材やニュース原稿などの放送素材を可視化するための放送素材可視化システムを試作した。このシステムでは、指定した単語の出現頻度の時系列変化を示すグラフや、共起する単語のワードクラウド、取材場所や取材対象の市町村を地図上にプロットするなど、色々な角度から可視化する。膨大な放送素材を俯瞰的に見ることのできるため、今まではなかなか気づけなかった取材対象の変化などを見つげられることが期待できる。

本稿では、システムの概要を紹介するとともに、2011年に発生した東日本大震災の放送素材を可視化した例を示す。

2. 放送素材アーカイブ

本稿では、分析対象として NHK 知財センターが保有している東日本大震災関連の放送素材アーカイブスのデータを利用した。今回は分析対象を以下のテキストデータのみとした。

- 約 30,000 件のニュース原稿
- 約 85,000 件の取材映像のキャプション¹

いずれも人手で書いた自然文のほか、作成された日時などの情報が付加されている。

今後、東日本大震災以外を扱った放送素材への拡張や、取材映像の音声認識による書き起こしの追加など、データの拡充を予定している。

¹ 映像の内容を把握できる程度の短い解説文や、インタビュウの書き起こしなどが含まれる。

[†] NHK 放送技術研究所
 NHK Science & Technology Research laboratories
[‡] NHK 知財センター
 NHK Rights & Archives Management Center

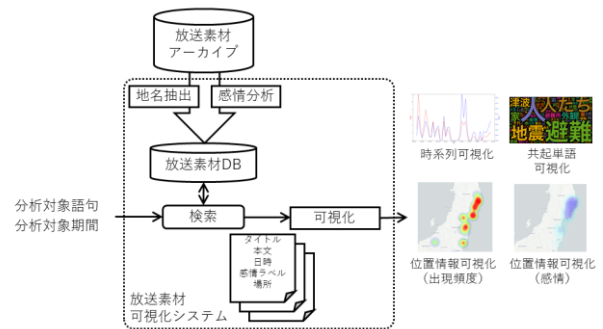
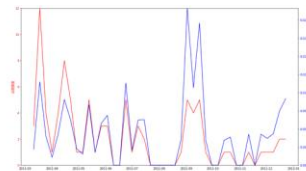


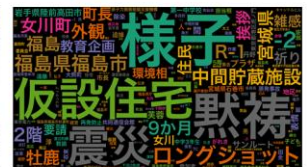
図 1 システム概要図



(a) 出現頻度の時系列変化



(b) 3/11-31 の共起単語



(c) 12/1-31 の共起単語

図 2 放送素材の時系列可視化例

3. 放送素材可視化システム

我々の開発した放送素材可視化システムは、ユーザからの分析対象語句や対象期間を入力として受け付け、分析対象語句が出現する放送素材について、複数種類の手法で可視化した結果を出力する。以下ではシステムの詳細について説明する。

3.1 システム概要

本システムの概要を図 1 に示す。本システムでは、まず放送素材アーカイブのデータをシステムに取り込む。その際、テキストに出現する地名の抽出と、テキストに現れる Positive / Negative の感情分析を行い、その結果をデータと合わせて DB 内に保持する。

可視化する際には、ユーザから受け付けた分析対象語句と対象期間を元に、DB から該当するデータを検索する。得られた検索結果を元に、複数の可視化を同時に行い、ユーザには簡単に見比べられるように可視化結果を提示する。

原稿内の地名の抽出には形態素解析エンジン mecab[3]を用いた。あらかじめ準備した都道府県-市町村をペアとし

た辞書を元に、市町村名と都道府県名が正しい組み合わせで出現している場合に、その市町村が原稿に出現したと扱う。また、感情分析には事前学習モデル BERT[4]の日本語版で感情分析に特化した fine-tuning を実施済みの bert-base-japanese-sentiment¹を使用した。

3.2 可視化手法

3.2.1 単語出現頻度の時系列変化の可視化

東日本大震災では、地震発生からしばらくの間、津波や原子力発電所による被害状況を伝えるための取材が中心であった。その後、被害状況から復興状況の話題に移り、さらに復興が進むことで明るみになった PTSD や農業・漁業への風評被害などに話題が変化していった。このように、時系列により話題は変化する。

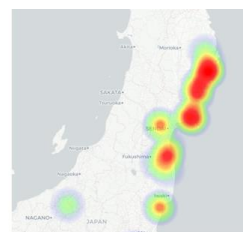
この変化を捉えるために、単語ごとの出現頻度の時系列変化を可視化した。図 2 に分析対象語句を「インタビュー」とした場合の例を示す。「インタビュー」という単語の出現頻度を可視化した図 2-(a) から、最も基本的な取材手法の一つであるインタビューでも、時期による出現頻度の変化があることがわかる。なお、図 2-(a) の赤線は分析対象語句が出現した素材数を、青線で表されるグラフは出現する素材が全体に占める割合をそれぞれ表す。

さらに、「インタビュー」と共起する単語を分析すると、震災が発生した直後である 3 月中 (図 2-(b)) の共起単語は「地震」「避難」などの震災そのものと関係の深い単語が多く出現するが、震災から期間を経た 12 月 (図 2-(c)) には「仮設住宅」「黙祷」など、3 月には出現しなかった単語が多く、「地震」や「津波」などの直接的に震災を表す単語の出現が大きく減っていることが読み取れる。このように、時系列の変化を可視化し、分析することで、取材対象の変化などが明確に読み取ることができる。原稿を 1 本ずつ見るのでは把握が難しい大きな流れを俯瞰するのに有用と考えられる。

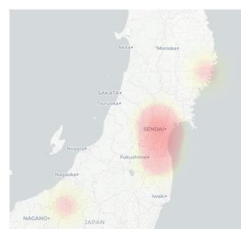
3.2.2 位置情報を用いた可視化

東日本大震災のように影響が広範囲に及んだ災害では、場所によっても話題が異なる。例えば岩手県の海沿いでの取材内容は津波に関するものが中心であったが、福島県の海沿い地域では、原発に関する取材が多い。一方で、岩手県での原発に関する取材や、福島県での津波に関する取材はそれほど数が多くない。このような、地域による取材の違いを可視化するために、放送素材に含まれる市町村名を用いて地図上への記事数のプロットによる可視化を行った。さらに、放送素材に含まれる文言から、Positive / Negative の感情極性判定を行い、その結果も地図上にプロットできるようにした。

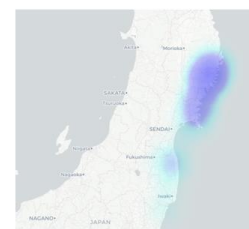
これらの可視化の結果を図 3 に示す。図 3 では、震災から約 2 ヶ月が経過した時点での「仮設住宅」という言葉が出現する放送素材の出現頻度、Positive / Negative な感情が表出している記事の出現頻度をそれぞれ市町村ごとに集計し可視化している。これにより、同じ時期の同じ言葉であっても、現れる感情が場所により大きく異なることがわかる。このような分布の違いを発見するためには、従来であ



(a) 出現頻度



(b) ポジティブな感情



(c) ネガティブな感情

図 3 放送素材の位置情報を用いた可視化例

れば人手で大量の放送素材を分析する必要があったが、本システムにより、その手間が大幅に削減できる。

なお、地図上へのプロットには folium²を、地図には OpenStreetMap³を使用した。

4. おわりに

放送局が保有している貴重な資料である過去の放送素材の有効活用を目的とし、放送素材可視化システムを開発した。本システムでは、放送素材を時系列と位置情報に基づいた可視化ができる。これにより、従来であれば人手で労力をかけていた過去の放送素材の分析が大幅に省力化できることから、放送素材の有効活用がしやすくなるものと考えられる。

今後は、感情分析の精度の測定などシステムの信頼性についての評価の実施や、分析対象とする放送素材の拡大、他の有用な可視化手法の検討などを進め、より使いやすいシステムを目指したい。

謝辞

放送素材可視化システムの Web インターフェースを開発してくださった、株式会社アスキングの松原 晋氏、甲田陽平氏に感謝いたします。

参考文献

- [1] 住吉 英樹, 三浦 菊佳, “言語特徴に基づくニュース記事分布の視覚化手法”, 映像情報メディア学会年次大会, 31B-1 (2017).
- [2] Philippe Laban, Marti Hearst, “newsLens: building and visualizing long-ranging news stories”, in Procs. Of the Events and Stories in the News Workshop, pp. 1-9 (2017).
- [3] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto, “Applying Conditional Random Fields to Japanese Morphological Analysis”, EMNLP 2004, pp. 230—237 (2004).
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, NAACL 2019, pp. 4171—4186 (2019).

¹ <https://huggingface.co/daigo/bert-base-japanese-sentiment>

² <https://python-visualization.github.io/folium/>

³ <https://www.openstreetmap.org/copyright>