

日本国内における原子力エネルギーに関する報道のトピックと感情分析

Analyzing Reports on Nuclear Energy in Japan Through Topic and Sentiment Analysis

*孫 一帆¹, 鶴田 博文², 熊谷 将也^{1,2}, 黒崎 健¹

¹京都大学, ²さくらインターネット株式会社

福島第一原子力発電所事故後、日本における原子力エネルギーの将来的な発展のためには、公共の感情を理解することが重要である。本研究では、原子力に関する YouTube 動画データを自然言語処理を用いて分析し、主要なトピックを特定し、関連するコメントに対して感情分析を実施した。

キーワード: 原子力, オンラインメディア, トピックモデリング, 感情分析, 言語モデル

1. 緒言

日本のエネルギーミックスにおいて重要な部分を占める原子力の円滑な発展には、政府と市民の効果的なコミュニケーションが必要である。これまで、世論調査は国民の感情を評価する効果的な方法であったが、用意された選択肢しか使えないことや、微妙な意見を捉えにくいという限界がある。そこで、本研究ではインタラクティブな YouTube の動画データとコメントを用いて、世論調査の補完的な手法の妥当性を探索する。

2. データ分析

2-1. データ収集

2024年4月1日までに、「福島第一原発」や「原子力発電」に関連する動画を対象として、日本の主要放送局の公式チャンネルから合計 3,084 本の動画を収集した。各動画のデータには、メタデータ (例: Like 数)、ビデオのタイトル、概要、コメント、および自動生成された字幕が含まれている。タイトル、概要、字幕のテキストを Sudachi[1]を用いて単語をトークン化し、全テキストデータで 30,290 のユニークな単語トークンの中から、具体的で意味のある名詞 4,957 語のみを選別した (図 1 に示している)。

2-2. トピックモデリング

各動画について、テキストの Bag-of-Words 表現を作成し、潜在ディリクレ配分法 (LDA) を用いてトピックモデリングを実施した。LDA により合計 11 のトピックが抽出され、各トピックに割り当てられた最も頻出する単語に基づいて、これら 11 のトピックが「処理水放出」「IAEA 視察」「地震・津波観測」「廃炉・デブリ処理」「使用済み核燃料・中間貯蔵施設建設」「最終処分場調査」「柏崎刈羽原発審査」「復興への思い」「賠償・費用」「原発稼働・運転期間」「原子力政策・電力」に関連していることが判明した。

2-3. 感情分析

各トピックについて、その関連コメントの感情を日本語評価極性辞書と大規模言語モデル (BERT と GPT) を用いて評価した。これらのモデルの性能についての詳細は学会にて発表する予定である。

3. 結論

本研究は、原子力分野における自然言語処理が大量のテキストデータを解析する有用性を強調し、日本における原子力に対する人々の感情を理解するための新しい分析アプローチを提案する。

参考文献

[1] Takaoka *et al.*, Sudachi: a Japanese Tokenizer for Business, 2018.

*Yifan Sun¹, Hirofumi Tsuruta², Masaya Kumagai^{1,2} and Ken Kurosaki¹

¹Kyoto Univ., ²SAKURA internet Inc.



図 1 作成したワードクラウド