

YouTube コメントに対するネットワーク分析に基づく 関心動向分析 ～環境問題を題材とした討論動画を対象として～

武内 奎太^{†1}, 松原 斎樹^{†1}, 岩崎 雅史^{†1}, 池田 維^{†2}

キーワード : YouTube, テキストマイニング, ネットワーク分析

1. はじめに

近年、動画共有サービスは世の中に広く浸透しており、代表的なサービスの1つである YouTube は若い世代を中心に多くの人から利用されている。YouTube にはコメント機能があり、視聴者はこの機能を通じて、動画の内容に対する感想や意見などを自由に投稿できる。語彙や文体の選び方、絵文字の使用のように表現の自由度が高く、投稿されたコメントには視聴者の率直な感情や本音が表れやすい。

YouTube 動画に対するコメント(以降、YouTube コメントと呼ぶ)を分析することは、アンケート調査などの従来の調査手法とは異なる視点から、人々の関心を読み解くのに有効な手段となり得る。しかしながら、YouTube コメントはテキストデータであるため、そのままでは定量的な分析に適さない。よって、YouTube コメントを効率的に分析するには、それらを数値化するなどの前処理も含め、系統的な分析手法を確立する必要がある。

本報では、環境問題を題材にした YouTube 動画[1]に対するコメントをもとに、視聴者の主要な関心を定量的に明らかにする方法を提案する。[1]は地球温暖化の実態をテーマに、環境問題に精通した2人の研究者と司会者による討論形式の番組である。[1]を選定した理由は、多くの視聴者によって活発な議論が展開されており、提案手法の有効性の検証に適していると判断したためである。

2. 方法

本節では、類似度の高い YouTube コメントの組合せを視覚化するための手法、それをもとに類似度の高い YouTube コメントで形成されるコミュニティを見い出すための手法、コミュニティを占めるポジティブ、ニュートラル、ネガティブなコメントの割合を算出するための手法を提案する。

まず、YouTube コメントを Google が提供する YouTube Data API[2]を用いて収集した。YouTube コメントは、動画に対するものと、それに対する返信の2種類に分けることができる。YouTube Data API のうち、前者に対しては「CommentThreads : list」を、後者に対しては「Comments :

list」を用いた。

次に、収集されたすべての YouTube コメントを、それらの意味を踏まえて数値ベクトルに変換した。この変換には、OpenAI が提供するテキスト埋め込みモデルである text-embedding-3-large[3]を用いた。

続いて、変換後のベクトルを用いて内容の類似度が高いコメントの組合せを特定し、コメントをノードに対応させた図式においてそれらのノードを線で結んだ。類似度評価には、コサイン類似度を用いてその値が 0.55 を超えるコメント同士を似ていると判定した。また、類似度が低い場合でも、親コメント(起点となるコメント)と子コメント(親コメントを受けてのコメント)は内容が共通している可能性が高いため、それらに対応するノード同士も線で結んだ。得られた図式は類似ネットワークと呼ぶことにする。

最後に、類似ネットワークに対して Clauset-Newman-Moore (CNM)法[4]を適用し、類似のノードが形成するコミュニティを抽出した。コミュニティ内の主要なトピックについては、コミュニティ内で隣接するノード数の多いコメントを読み解くことで明らかにした。また、Google が提供する Cloud Natural Language API[5]による感情分析を用いて、各コメントの感情をポジティブ、ネガティブとどちらにも属さないニュートラルのいずれに基づくものかを判別し、それらが占める割合をコミュニティごとに算出した。Cloud Natural Language API では、入力されたテキストに対する感情を-1~1 の値で評価する。本報では、収集した YouTube コメントに基づいて、0.15 以上をポジティブ、-0.15 以下をネガティブ、それ以外をニュートラルと判別した。

3. 分析結果

[1] の YouTube コメント 4,407 件に対して類似度分析を行うと図 1 のような類似ネットワークが得られた。続けて、CNM 法[4]を適用すると、184 個のコミュニティが抽出された。本報では、その中でも多くの視聴者の関心が反映されていると考えられる、コメント数上位 5 つのコミュニティに着目した。それぞれのコミュニティを、コメント数が多い順に C_1, C_2, C_3, C_4, C_5 と記すことにする。

^{†1} 京都府立大学

^{†2} 産業技術総合研究所

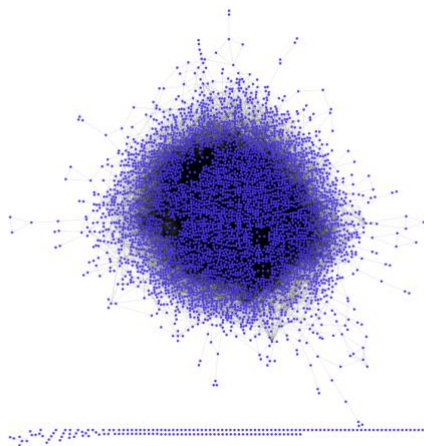


図 1 [1]の YouTube コメント 4,407 件より得られた類似ネットワーク

表 1 $C_1 \sim C_5$ に含まれるコメント数と主要なトピック

| | コメント数 | 主要なトピック |
|-------|-------|--------------------------|
| C_1 | 1,077 | 地球温暖化の原因 |
| C_2 | 1,056 | 地球温暖化の科学的根拠とそれを支える研究について |
| C_3 | 934 | 他者のコメントに対する挑発的な発言 |
| C_4 | 482 | 出演者の発言に対する意見 |
| C_5 | 224 | 地球温暖化に対する世界の動向 |

表 2 $C_1 \sim C_5$ を形成するコメントのポジティブ/ニュートラル/ネガティブの判定結果

| | ポジティブ | ニュートラル | ネガティブ |
|-------|-------|--------|-------|
| C_1 | 3.5% | 64.5% | 32.0% |
| C_2 | 3.7% | 42.4% | 53.9% |
| C_3 | 5.0% | 40.5% | 54.5% |
| C_4 | 10.2% | 37.3% | 52.5% |
| C_5 | 5.4% | 46.4% | 48.2% |

まず, $C_1 \sim C_5$ のコメント数と主要なトピックを表 1 に示す. 表 1 より, コメント数が特に多いのは C_1, C_2, C_3 であることが分かる. また, C_1, C_2, C_5 では地球温暖化, C_3 ではコメントに対する挑発的な発言, C_4 では出演者に関するトピックが支配的である. 次に, $C_1 \sim C_5$ を形成するコメントのポジティブ/ニュートラル/ネガティブの判定結果を表 2 に示す. 表 2 より, C_1 はニュートラル, C_2, C_3, C_4, C_5 はネガティブに判定されたコメントが最も多くを占めていることが分かる. また, ニュートラルを除いたポジティブ/ネガティブの判定結果に着目すると, すべてのコミュニティでネガティブ判定の割合が圧倒的に多く, 否定的な言葉を用いたコメントが優勢であることが読み取れる.

4. まとめと今後の展望

本報では, ネットワーク分析に基づいて[1]に対するコメントをそれらの内容に応じた類似度をもとにコミュニティに分類し, 視聴者の主要な関心を定量的に明らかにした. ただし, 当然ながら YouTube コメントのみを対象とした分析では, YouTube を利用しない層の関心を捉えることはできない. 今後の研究課題は, X や Instagram などの SNS 上の投稿も分析の対象とすることで, より多様な人々の関心が把握できるような分析手法を提案することである.

参考文献

- [1] “【激論】IPCC 報告書の執筆者 2 人が対立「地球温暖化は本当か？」”. <https://www.youtube.com/watch?v=gig8VkcLeUo>, (2026 年 1 月 10 日アクセス).
- [2] “YouTube Data API Reference”. <https://developers.google.com/youtube/v3/docs?hl=ja>, (2026 年 1 月 10 日アクセス).
- [3] “text-embedding-3-large Model | OpenAI API”. <https://platform.openai.com/docs/models/text-embedding-3-large>, (2026 年 1 月 10 日アクセス).
- [4] A. Clauset, M.E.J. Newman, and C. Moore, “Finding Community Structure in Very Large Networks,” *Physical Review E*, Vol. 70, No. 6, 066111 (2004).
- [5] “感情分析 | Cloud Natural Language API”. <https://docs.cloud.google.com/natural-language/docs/analyzing-sentiment?hl=ja>, (2026 年 1 月 13 日アクセス).