

動画共有 SNS のコメント分析と視聴者行動ネットワーク分析による 金融コミュニティの構造抽出

鈴木 隆広^{1,2} 永野 清仁²

¹株式会社群馬銀行

²群馬大学大学院情報学研究科

{j241a602, k-nagano}@gunma-u. ac. jp.

概要

近年金融業界において、金融テキストマイニングへの関心が高まっており、既存研究では SNS データや BERT を用いた金融テキストの感情分析などが行われている。本研究では、動画共有 SNS における金融商品関連のコメントデータについて、ユーザーコメントを金融カテゴリに分類する BERT モデルと、ユーザー行動データに基づくコミュニティ分析の結果を組み合わせることで、金融商品に対する世間の関心およびその構造の把握を試みる。

1 はじめに

近年、金融業界においては、金融テキストマイニングへの関心が高まっている [1,3]。これまでに、SNS の投稿データと株式市場の動向の関連性を検証する研究 [2] や、景気ウォッチャー調査を対象とした感情分析 [3] 等が報告されている。

金融関連テキストの分析で近年よく用いられる技術として BERT [4] がある。BERT により、事前学習済みモデルと少量のラベル付きデータによるファインチューニングを通じて、高い精度のデータ分類が可能となる。BERT を用いた金融テキストを対象とした分析の多くは、筆者の知る限りでは、感情の二値・三値分類などが中心的な話題である。

SNS データは、ユーザーをノード(点)、関係性をエッジ(線)で表すことで、ネットワーク科学の分析対象として捉えることができる。SNS ユーザーにより形成されるネットワーク構造の分析は、対象の理解の手助けとなる [5,6]。金融商品に興味を持つユーザーを対象としたネットワーク分析は、金融情報に対する世間の関心や特性の深い理解につながる事が期待できる。

本研究では、動画共有 SNS である YouTube の金融商品関連動画のコメントデータを対象とする。YouTube では、動画配信者が登録した「チャンネル」ごとに複数の動画がまとめられており、ユーザーは(一部動画を除き)各動画へのコメントが可能である。このため、ユーザー間だけでなく、チャンネル間の関係性も分析対象となる。本研究では、ユーザーコメントを金融商品に関するマルチラベルに分類する BERT モデルと、ユーザー行動データに基づくユーザーコミュニティ・チャンネルコミュニティの抽出結果を組み合わせることで、金融商品への SNS ユーザーの関心とその構造の把握を試みる。

2 データ取得と BERT モデル

本章では、YouTube における金融関連データ(動画、コメント、ユーザー、チャンネル等のデータ)の取得方法と、YouTube コメント分析のための BERT モデルの作成について述べる。

2.1 データの取得

本研究では金融テキストとして YouTube における金融関連動画へのコメントを用いる。YouTube は国内・国外で高い利用率を有する動画共有 SNS の一つであり [7]、API を用いて動画コメントを取得可能である。

金融カテゴリとチャンネルデータについて。 金融商品に関するカテゴリとして「保険」「資産運用(株式・投資信託など)」「不動産」「相続」の4分野を設定した。分析対象とする YouTube の金融関連チャンネルの絞り込みのために、キーワードとして、資産運用に関連する 20 語、保険、不動産、相続それぞれに関連する各 10 語、計 50 語の金融用語を独自

に設定した。各キーワードの関連動画について、視聴回数上位の 500 件からショート動画を除外した上でチャンネル情報を抽出し、4 つの金融カテゴリごとにチャンネル登録者が多い順に 30 件を抽出した（ただし、討論形式の動画が中心のチャンネルおよび証券会社等の公式チャンネルなどは分析対象から除外）。さらに重複を除外することで、対象とするチャンネル数は 99 件となった。

動画・コメントデータについて。 得られたチャンネルに投稿されている動画のうち、2024 年 4 月 1 日から 2025 年 3 月 31 日までの期間の投稿されたものを分析対象とする。金融関連でない動画を除外するために、動画タイトルから金融関連動画のフィルタリングし、結果として動画 5,462 件を取得した。フィルタリングでは、証券会社（みずほ・大和）が公開する金融用語辞書を基本とし、新しい用語（例えば新 NISA など）に関する独自のユーザー辞書を合わせて用いた。続いて、抽出した動画の全コメントを取得し、同様に金融用語辞書によるフィルタリングを実施し金融コメントを取得した。

2.2 金融商品 BERT モデル

2.1 節で抽出した金融コメントのうち金融カテゴリ 4 分野から同一割合となるようコメントをランダムに抽出、計 2,000 件をアノテーション対象データとし、著者自身によりマルチラベル（不明を含む）を付与した。その結果が表 2.1 の通りである。

表 2.1. 付与したマルチラベル (2000 件)

マルチラベル	コメント数
資産運用	562 件
不明	241 件
相続	207 件
資産運用, 不動産	189 件
不動産	189 件
資産運用, 相続,	173 件
保険, 資産運用	172 件
保険	104 件
不動産, 相続	52 件
資産運用, 不動産, 相続	51 件
その他	60 件

東北大学乾研究室が公開している事前学習済み日本語 BERT モデル [cl-tohoku/bert-base-japanese-v2](https://github.com/ku-nlp/cl-tohoku/bert-base-japanese-v2)

を使用し、学習用データと検証用データの割合を 9:1 としてホールドアウト検証による分類モデルの評価を行った。その結果、完全一致率は 75%、ハミング損失は 5.2%となった。

このように得られた分類モデルを以下、金融商品 BERT とよぶ。金融商品 BERT は、各金融コメントを受け取ったとき、保険、資産運用、不動産、相続、不明の 5 つの項目について 0 以上 1 以下の値を返す。

3 コメント内容に基づく分類

2.1 節で取得した金融コメントについて、コメント投稿者（ユーザー）のうち、10 件以上コメント投稿しているユーザーのみを以下の分析対象とする。この抽出の結果、分析対象のチャンネル数は 98、ユーザー数は 2,225、コメント数は 44,220 件となった。

金融商品 BERT を用いて、2,225 人の各ユーザーに金融（マルチ）ラベルを以下の手順で付与する。

ユーザーへの金融ラベル付与：

ユーザー A が投稿した分析対象コメントが n 件について、金融商品 BERT が返す 5 項目（保険、資産運用、不動産、相続、不明）のスコアを算出し、1 つ以上のコメントで最大スコアかつそのスコアが 0.9 以上となる項目を、「不明」を除外して、集めたものをユーザー A の金融ラベルとする。このような項目が 1 つもないユーザーは未付与として扱う。

上記の金融ラベル付与をユーザー 2,225 人に適用した結果を表 3.1 に示す。

表 3.1. ユーザーに付与された金融ラベル

金融ラベル	ユーザー数
資産運用	763 人
資産運用, 保険	419 人
不動産, 資産運用, 保険	348 人
不動産, 資産運用	329 人
相続, 不動産, 資産運用, 保険	131 人
相続, 資産運用, 保険	96 人
相続, 資産運用	74 人
相続, 不動産,	54 人
その他（未付与含む）	11 人

ユーザーの場合と同様に、各チャンネルについても動画へのコメントに対して金融商品 BERT を適用して金融（マルチ）ラベルの付与する。ただし、ユーザーの場合「1 つ以上のコメントで」としたところ

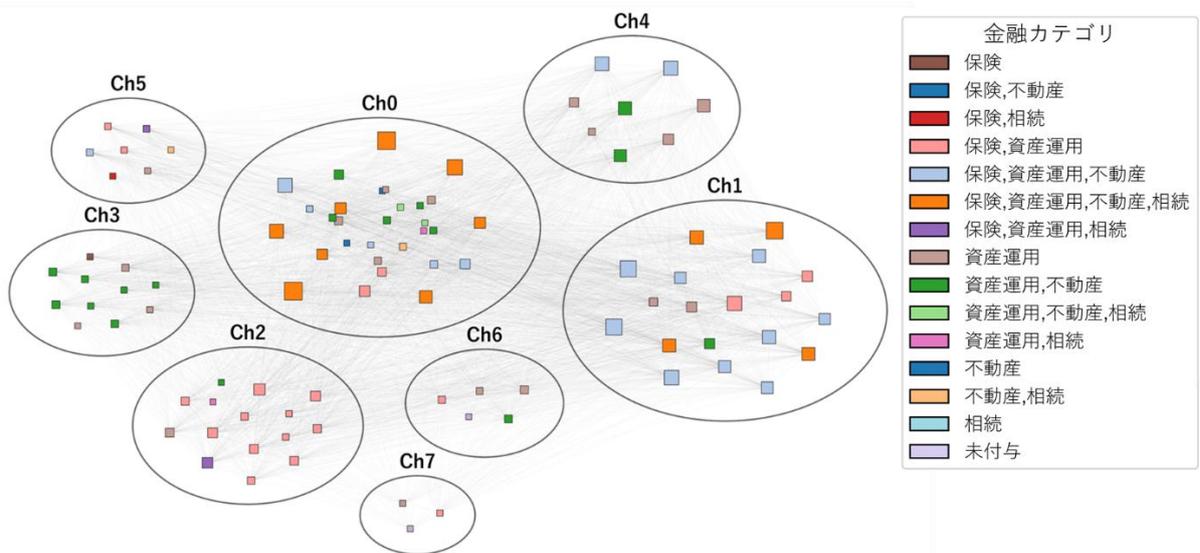


図 4.1. YouTube における金融関連チャンネルのコミュニティ構造

を、チャンネルの場合「10%以上のコメントで」と変更して金融ラベルとして付与を行った。この手順をチャンネル 98 個に適用した結果を表 3.2 に示す。

表 3.2. チャンネルに付与された金融ラベル

金融ラベル	チャンネル数
保険,資産運用	20
資産運用	19
不動産,保険,資産運用	17
不動産,資産運用	17
不動産,相続,資産運用,相続	12
その他 (未付与含む)	13

4 ネットワーク分析に基づく分類と金融商品 BERT モデルの融合

3 章と同じ分析対象について、4.1 節ではまずコメントの内容ではなく、コメントしたユーザーや動画の「重複度合い」でユーザー間やチャンネル間の類似度を定め、その上でネットワーク科学分野のコミュニティ検出手法を適用することを考える。ここではダイス係数をベースとした値を用いる。加えてネットワーク分析で得られたコミュニティ構造に 3 章で得られた BERT を用いた金融ラベル情報を組み合わせることで、チャンネル構造 (やユーザー構造) を多面的にとらえることのできる分析手法を提案する。また、4.2 節ではユーザーコミュニティとチャンネルコミュニティの成す構造を二部グラフとしてと

らえて分析手法を提案するとともに、BERT により得られる金融ラベル情報を組み合わせた分析のアプローチを提案する。

4.1 チャンネル間・ユーザー間の分析

チャンネル間のネットワーク分析に基づく分類について説明する。各チャンネルをノードとみなし、チャンネル X と Y の類似度 $\text{sim}_{\text{ch}}(X, Y)$ として次式で定める。

$$\text{sim}_{\text{ch}}(X, Y) = \frac{2 |W_X \cap W_Y|}{|W_X| + |W_Y|}$$

ここで W_X, W_Y はそれぞれチャンネル X, Y にコメントしたユーザー集合であり、共通するユーザーが多いチャンネル同士ほど類似度は大きくなる。ノード間のエッジの重みは、正のパラメータ λ を用いて、 $(\text{sim}_{\text{ch}})^\lambda$ の形で与える。このように構築したネットワークについて、コミュニティ検出手法 (Newman のモジュラリティに基づく手法など) を適用し、チャンネル全体のコミュニティ構造が得られる。同様の考え方でユーザー間についても、コメントした動画の重複度合いに基づいた類似度を設定することで、ユーザー全体のコミュニティ構造が得られる。

ネットワーク分析で得られたコミュニティ構造と金融商品 BERT によって 3 章で付与したチャンネルへの金融ラベルを組み合わせる結果を図 4.1 に示す。このように SNS データについて複数の視点からの分析結果を組み合わせることによって、対象への深い理解につながることを期待できる。

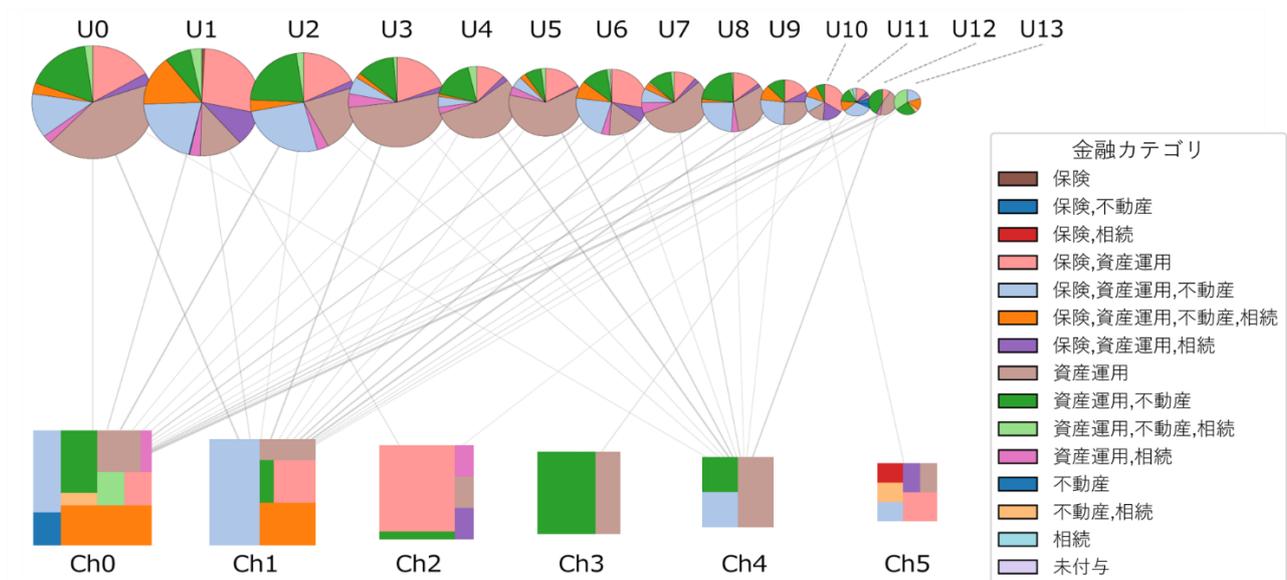


図 4.2. YouTube における金融関連チャンネルコミュニティとユーザーコミュニティの二部グラフ構造

4.2 二部グラフによる全体構造の分析

ユーザー集合とチャンネル集合を合わせた全体の構造を分析するにあたり、各々の要素をそのまま分析するのでは構造が複雑になる。4.1 節で示したチャンネルコミュニティ、ユーザーコミュニティの単位で粗くとらえた二部グラフとして扱うことで見通しのよい分析を期待できる。また、金融商品 BERT によるラベルを組み合わせた分析も可能である。図 4.2 は分析データに関するチャンネルコミュニティとユーザーコミュニティにより形成された二部グラフ構造を示している。Ch0, ..., Ch5 がチャンネルコミュニティ、U0, ..., U13 がユーザーコミュニティを表すノードであり、各コミュニティノードには金融ラベルの割合情報が付与されている。またコミュニティ間のエッジ情報は、ユーザーコミュニティ側のチャンネルコミュニティ内動画への平均コメント数をベースとして表示している。

4.3 考察

各チャンネルコミュニティの特性。 図 4.1, 図 4.2 の各チャンネルコミュニティについて、登録者数上位 3~5 件のチャンネルを確認したところ各コミュニティの以下の特性が確認できた。

- Ch0 : 専門家による知識提供型のコミュニティ
- Ch1 : 初心者向け金融教育コミュニティ
- Ch2 : 生活に直結した情報が多い

Ch3 : 投資初心者向け、不動産が中心の特化型

Ch4 : 株式・相場分析に特化、中上級者向け

Ch5 : 高齢層・富裕層・経営者向けの専門特化

Ch6 : 有名人・エンタメ要素と金融知識が混在

Ch7 : ポイ活・節約などニッチ領域

金融商品 BERT によるマルチラベルが混在するチャンネルコミュニティが多いが、実際のコミュニティごとのチャンネルには明確な特性のもと形成されることが確認された。

チャンネル特性とユーザー特性。 Ch0, Ch1 はあらゆる金融カテゴリを扱っており、図 4.2 から幅広いユーザーに視聴されていることがわかる。また、Ch2 と Ch3 のノード情報からテーマの偏ったチャンネル群であるが、接続するユーザー情報から特定の分野に強く興味を持つユーザーの存在が確認できる。個々のデータから把握が困難な特性もネットワークを構築することで理解が可能であることが示される。

5 おわりに

本研究では、YouTube 上の金融関連コメントを対象として、BERT によるコメント分析とネットワーク分析によるコミュニティを形成し、また両手法を組み合わせた分析手法を提案した。今後の展望として、コメント分析とネットワーク分析を統合により得られる知見を活かすことで、顧客に金融商品を推薦する仕組みへの応用が課題の 1 つである。

参考文献

- [1] 高野海斗, 中川慧, 藤本悠吾. 大規模言語モデルを用いた金融テキスト二値分類タスクの定義文生成とチューニング手法の提案. 人工知能学会研究会資料, ISSN 2436-5556, 2024.
- [2] Johan Bollen, Huina Mao, Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*. arXiv preprint arXiv:1010.3003, pp. 1-8, 2011.
- [3] 高野海斗. 金融テキストを対象とした強弱を捉えることができるセンチメントモデル開発のためのデータセット構築方法の検討および分析. 人工知能学会第二種研究会資料, No. FIN-034, pp. 98-108, 2023.
- [4] 今井康太, 酒井浩之, 永並健吾, 稲垣真太郎. 債券市場における金融極性辞書の自動構築とその拡張. 人工知能学会第二種研究会資料, No. FIN-028, pp. 171-, 2022.
- [5] 大竹恒一, 生田目崇. ソーシャルメディア情報-を用いた消費者コミュニティの理解-ファッションブランドにおけるコミュニティ構造分析-. 日本オペレーションズ・リサーチ学会, ISSN 0030-3674, pp. 498-504, 2022.
- [6] 松尾豊, 安田雪. SNS における関係形成原理—mixi のデータ分析—. 人工知能学会, pp. 531-541, 2007.
- [7] 株式会社ホットリンク. 2025 年 | 日本・世界の SNS 利用者数ランキングまとめ! SNS 別のマーケティング成功事例も解説. 2025. 1. 6 (参照日: 2026. 1. 6)
https://www.hottolink.co.jp/column/20250106_114872/.
- [8] 五十嵐祐, 笹原和俊. 日本語テキスト解析のための感情辞書標準版の作成と公開: ビッグデータ時代の心理分析に向けて. 電気通信普及財団研究調査助成報告書, 2018.
- [9] 五島圭一, 新谷元嗣, 高村大也. 景気単語極性辞書の構築とその応用. 言語処理学会, pp. 1233-1253, 2022.