

<論文>

趣味としての超小型衛星開発コミュニティの長期的変容 —チャットログのテキストマイニングによる分析—

Long-Term Transformation of a Hobby Nanosatellite Development Community: Text Mining Analysis of Chat Logs

渡辺謙仁*

Takahito Watanabe

I. はじめに

今や宇宙開発は、かつてのように一部の大国のみによって行われている訳ではない。「キューブサット」と呼ばれる、10 cm 立方の手の平サイズの超小型衛星の規格が存在することをご存じだろうか。このキューブサットが 1999 年に仕様策定され、2003 年に東京大学が開発した衛星などが打ち上げられて以来、衛星の開発と打ち上げの経済的・技術的障壁が大幅に下がった。その影響は学术界や産業界のみならず趣味縁にまで及び、日本国内においても（世界的にも先駆的な試みではあるのだが）、2018 年にはリーマンサット・プロジェクト（RSP）が開発した衛星が宇宙空間へ放出されるに至った。本稿では、RSP のような超小型衛星開発の趣味縁の日本における原型と思われる「ソーシャルメディア衛星開発プロジェクト SOMESAT」を取り上げ、そこで使用されていたチャットのログのテキストマイニングにより、プロジェクトの長期的変容を分析する。

II. 方法

1. 対象の概要

本稿で分析の対象とする「ソーシャルメディア衛星開発プロジェクト SOMESAT」(Social Media Satellite Development Project。以下、SOMESAT) (図 1) とは、バーチャルシンガー初音ミクの派生キャラクターである「はちゅねミク」を乗せたキューブサットを開発して宇宙へ打ち上げ、宇宙でミクにネギを振らせることを目的として始まったプロジェクトである。同様のプロジェクトは 2009 年の初め頃、HAXA: Hachune Aerospace eXploration Agency (はちゅね宇宙航空研究開発機構) や CMS: Character Media Satellite (キャラクターメディア衛星) などと呼ばれ、基礎的な実験やニコニコ動画上での情報共有、JAXA に対する説明などが行われていた。その後、ニコニコ動画上でものづくりの動画を個人が自由に投稿する活動であるニコニコ技術部のオフ会である「NT 京都」(2009 年 3 月 21 日開催)での口頭プレゼンを経て、SF 作家の野尻抱介が提案する CMS が、技術者である超電磁 P (「超電磁 P」はハンドルネーム) が提案する HAXA に合流し、IRC: Internet Relay Chat と呼ばれるチャット上で、プロジェクトに関する議論が行われるようになる。本稿で分析の対象にするのは、この IRC のログである。HAXA と後の SOMESAT で

* 静岡大学グローバル共創科学部

使用していた IRC は、当時からニコニコ技術部の活動をしていた「ニコニコ技術部員」の1人がニコニコ技術部の情報交換のために構築したものであり、ユーザー登録は必要なく誰でも書き込むことができた。またログも誰でも閲覧可能であったが、現在は閉鎖されログは見られなくなっている。なお、現在の SOMESAT ではログがより確実に残り、同期的コミュニケーションの必要性から解放されるといった理由から、IRC の代わりに Slack が使用されている。



図 1. 2010 年 5 月頃、イベントでの配布用に作成されたチラシの一部 (筆者提供)

HAXA は後に、HAXA という名称では JAXA をもじっており、JAXA と衛星打ち上げ契約を結ぶことになった際によく思われないであろうこと、また当時は初音ミクをはじめとした歌声合成ソフトウェア「ボーカロイド」はまだサブカルチャーの域を出ていなかったこともあり、ミクのネギ振りというニコニコ技術部の内輪の「ノリ」を世間一般に晒すようなことはみつともないなどといった意識から、CMS に似た SOMESAT に名称変更することになる。ニコニコ技術部のような「メイカーズムーブメント」ないし「ソーシャルアプリケーション」、つまり工業の個人化と、それを巡る趣味を介した社会的なつながりや、初音ミクと N 次創作、キューブサットなどの宇宙開発の社会技術史に関しては本稿の範疇を超えるので、参考文献(渡辺, 2015; 渡辺・田邊, 2016)を参考にしたい。

2. データの収集方法

既に述べたように、本稿で分析する IRC のログは誰でも閲覧可能であったものである。連番ダウンロードソフトウェアを用いて 2009～2016 年に記録された IRC のログを取得し、ファイル結合ソフトウェアを用いてダウンロードした IRC の全てのログを 1 つのファイルに結合した。この段階では、ログは PHP で作成されたチャットロガーによって生成された、HTML タグ形式のテキストファイルになっていた。このテキストファイルを、Python を用いたコーディングによ

って CSV 形式のデータに変換した。

3. 分析手法

上記の CSV 形式のデータを、テキストマイニングソフトウェアである KH Coder (樋口, 2020) や Excel のピボットテーブルを使用し、分析した。行った分析は、共起ネットワーク分析や対応分析 (KH Coder)、ハンドルネームやキーワードの出現数の経年変化など (Excel) である。

データの分析に当たっては、周ほか(2024)を参考にした。従来型統計を用いた計量テキスト分析は、プロセスの明瞭性と検証可能性を備えた信頼できる内容分析の方法である一方で、その結果解釈には分析者に一定の統計リテラシーと想像力が求められるため、解釈の質が分析者の力量に大きく左右されるという課題がある(周ほか, 2024)。そこで本研究では、KH Coder を用いた計量分析に対し、生成 AI の一つである LLM: Large Language Model (大規模言語モデル) を補完的に活用することで、この課題に対応することにした(周ほか, 2024)。KH Coder の再現可能な分析結果を人間が確認しつつ、LLM の出力を一つのヒントとして活用すれば、結果解釈の信頼性と質の底上げに寄与できることが期待された(周ほか, 2024)。以下、KH Coder と LLM を用いた分析手法について詳しく述べる。

(1) データ

上記の CSV 形式のデータを使用した。また、対応分析に用いる外部変数として、「年」(2009～2016年)を採用した。

(2) 使用した KH Coder

分析に用いた KH Coder のバージョンは、3.Beta.05 であった。また、前処理時の形態素解析には、「ChaSen」を指定した。

(3) LLM の選定

分析に用いる LLM として、OpenAI 社の ChatGPT-o3 (以下、ChatGPT) を選定した。なぜなら、今回の分析ではテキストデータのみならず、KH Coder から出力された分析結果の画像を解析するために、マルチモーダル機能が必要であったためである(周ほか, 2024)。ChatGPT には 4o からマルチモーダル機能が備わっているが、o3 ではその機能が向上している。

(4) KH Coder による分析結果の出力

データを KH Coder に読み込み、前処理を行った後、抽出語を対象とした「共起ネットワーク分析」、ならびに「対応分析」を行った。その後、分析で得られた結果を PNG 形式のカラー画像として保存した。

(5) 分析結果解釈用プロンプトの作成

KH Coder の分析で得られた結果画像を ChatGPT にアップロードし、そこから解釈のヒントを得るためのプロンプトを ChatGPT に与えた。プロンプトの作成に当たっては周ほか(2024)を参考にした。

4. 倫理的配慮

結果の図表を本稿に掲載する際、ハンドルネームは伏せた。また、本稿の投稿前に、真実性やプライバシーなどが確保されているかどうか、SOMESAT のメンバーに確認を求めた。

##背景### あなたは「テキストデータ分析の専門家」です。IRC (Internet Relay Chat) のログ (テキストデータ) に対して、統計的な手法を用いて分析された結果を考察することができます。今回取り上げるIRCは、ニコニコ動画でものづくりの様子を見せ合う活動であるニコニコ技術部を母体とし、初音ミクの派生キャラクター「はちゅねミク」を超小型衛星CubeSatに乗せ、宇宙でネギを振らせることを目的として結成されたプロジェクトである「SOMESAT (SOcial MEdia SATellite development project)」とその前身である「HAXA (Hachune Aerospace Exploration Agency)」で使われていたものです。この度、このIRCのログから言葉を抽出し、言葉の関係性を表す共起ネットワーク図を作成しました。この共起ネットワーク図を基に、IRCのログに書かれているテーマを分析することになりました。

###指令### 以下の「###共起ネットワークの定義###」と「###サブグラフの定義###」を理解し、私が提示した共起ネットワーク図に含まれるすべてのサブグラフを解釈し命名してください。その際に、以下のステップを辿ってください。

###サブグラフの解釈と命名の例### <<サブグラフ 01>> サブグラフに含まれる言葉: "IRC", "ログ", "確認" 解釈: IRCのログを確認したことに関する話題。
テーマ名: ログ確認

###共起ネットワークの定義### 共起ネットワークは、テキストデータの中で特定の単語が同時に出現する (共起する) 関係をネットワークとして表現したものです。ノード (頂点) が単語を表し、エッジ (辺) がそれらの共起関係を表します。

###サブグラフの定義### 共起ネットワークにおいて、サブグラフは元の共起ネットワーク (全体グラフ) からエッジによって繋がっているノードを取り出して構成されるネットワークです。エッジで繋がっていないノードは同じサブグラフに属しません。

図3. サブグラフの解釈プロンプト (筆者作成)

No.	テーマ名	含まれる言葉	解釈
1	プロジェクト参加・メンバーシップ	「SOMESAT」「参加」「人」	プロジェクト本体 SOMESAT に「参加」する「人（メンバー）」をめぐる発言がまとまっています。メンバー募集・自己紹介・役割分担決めなど、人の出入りや参加表明をやり取りしている場面と読めます。
2	個人情報の扱い	「情報」「個人」	個人を特定する／しない情報の取り扱い、あるいはプライバシーと公開情報の区別に関する短いやり取りが収束しています。
3	作業終わりの挨拶・離席宣言	「疲れ」「放送」「今日」「失礼」「おやすみなさい」「落ちる」	深夜帯や長時間作業後の「疲れ」を共有し、「今日はここまで」「落ちる（離席する）」「失礼」「おやすみなさい」といったログアウト前のあいさつが多い雑談クラスターです。生放送（ニコニコ生放送など）終了の合図も混じっています。
4	ログ確認・ログ参照	「IRC」「ログ」「確認」	IRC 上の発言ログを「確認」したり、ログを参照して話を進めたりするメタコミュニケーション。ログ保管／読み返しに関する話題です。
5	業務連絡・進捗報告	「報告」「事務」「連絡」「お願い」「M()m」「学生」「技術」「他」「雑談」「無い」「ありがとう」	プロジェクトの「報告」や「事務」連絡、作業「お願い」、技術的な進捗共有など、日次・週次の業務連絡チャンネルに相当するやり取りが集中しています。丁寧語や謝辞（M()m, ありがとう）が多く、公式な連絡・報告フローを示唆します。
6	衛星開発・宇宙技術	「開発」「衛星」「宇宙」	キーワードがダイレクトに「衛星」「宇宙」へ「開発」する技術的コアトピック。ハードウェアやミッション設計そのものについて議論している最小クラスターです。
7	リンク貼り付け／HTML タグ断片	「href=」「target=」「_blank」「(a)」「NT」	HTML アンカータグの属性断片が固まっており、IRC ログにコピペされたリンクコード（ など）がそのままトークン化されたものと考えられます。

表 1. サブグラフの解釈に関する応答（ChatGPT の応答を元に筆者作成）

あなたが提案したテーマを検討した結果、以下の7個のテーマに注目することにしました。

テーマ1：プロジェクト参加・メンバーシップ

テーマ2：個人情報の扱い

テーマ3：作業終わりの挨拶・離席宣言

テーマ4：ログ確認・ログ参照

テーマ5：業務連絡・進捗報告

テーマ6：衛星開発・宇宙技術

テーマ7：リンク貼り付け／HTMLタグ断片

続いては、原文を参考に、共起ネットワーク全体の考察文を考えてください。その際に、以下の「ポイント」を考慮してください。原文は添付したCSVファイルのG列にあります。

ポイント1：テーマ1からテーマ7のすべてのテーマを考察すること

ポイント2：考察の根拠を原文を引用して説明すること

図4. 共起ネットワーク全体の考察文生成プロンプト（筆者作成）

総合的な示唆

1. サブグラフ 05 (業務連絡) と 06 (衛星開発) がネットワークの二大極。前者がプロジェクト運営を、後者が技術的核心を担い、両者の橋渡し役として 04 (ログ確認) が分散接続している。
2. 03 (挨拶) は全体をゆるく横断し、疲労共有 → 離席宣言 → 次のタスクへというリズムを生む。
3. 01 (メンバーシップ) と 02 (個人情報) はコミュニティの拡大と成熟を示す対概念。新規参加が増えるほど個人情報保護の議論が顕在化している。
4. 07 (リンク貼り付け) は自動生成ノイズでありつつ、外部成果物へのアクセスポイントとして機能し、オープンコラボレーションの実態を裏付ける。

以上より、SOMESAT/HAXA の IRC では「遊び心を核に据えた本気の宇宙開発」が、業務連絡と技術議論を往復しつつ、メンバー参加・情報共有・プライバシー配慮という三つの社会的課題を並行処理する形で進行していることが、共起ネットワークから立体的に把握できる。

図 5. 共起ネットワーク全体の考察文の応答 (一部) (ChatGPT の回答)

ChatGPT は 7 つのサブグラフを正しく抽出している。また、表 1 のサブグラフに「含まれる言葉」はすべて、図 2 の共起ネットワークのサブグラフに含まれる言葉に正しく対応しており、テーマ名や解釈にもおかしなところはない。一方で、図 5 の共起ネットワーク全体の考察文の応答を見ると、「サブグラフ 05 (業務連絡) と 06 (衛星開発) がネットワークの二大極。前者がプロジェクト運営を、後者が技術的核心を担い、両者の橋渡し役として 04 (ログ確認) が分散接続」といった記述や、「03 (挨拶) は全体をゆるく横断」といった記述は、実際の共起ネットワークとは異なったものになっている。

2. 対応分析

KH Coder の分析で得られた対応分析の結果をもとに、ChatGPT に対応分析の解釈方法に従って結果解釈させた。分析では、KH Coder の「抽出語」「対応分析」メニューを選択し、外部変数を「年」に指定した。この結果 (図 6) を PNG 形式のカラー画像ファイルとして保存した。

その後、対応分析の結果画像 (図 6) を ChatGPT にアップロードし、結果の解釈文を提案させるプロンプト (図 7) を実行した。それに対する ChatGPT の応答は表 2 や図 8 のようになった。引用例の出力ではハンドルネームも示されていたが、表 2 では伏せた。図 8 の応答を見る限り、文脈に沿わない、おかしなところはない。一方で表 2 で示した応答では、原点の近くにある語や、原点から変数 (赤い四角) に向かって 10° の角度の範囲に収まらない語を特徴語 (代表語) として挙げてしまったり、無線という文脈に沿わせようとして「WWV」(アメリカ合衆国コロラド州フォートコリンズにある、アメリカ国立標準技術研究所(NIST)が管轄していた短波無線局の意か) という対応分析の図中には存在しない語を特徴語として挙げてしまったり、ログ確認という文脈で、「USB2.0 で...動作確認とれている...?」といった文脈に沿わない引用例を挙げてしまったりしたが、全体的な特徴は大きく外さずに掴めている。

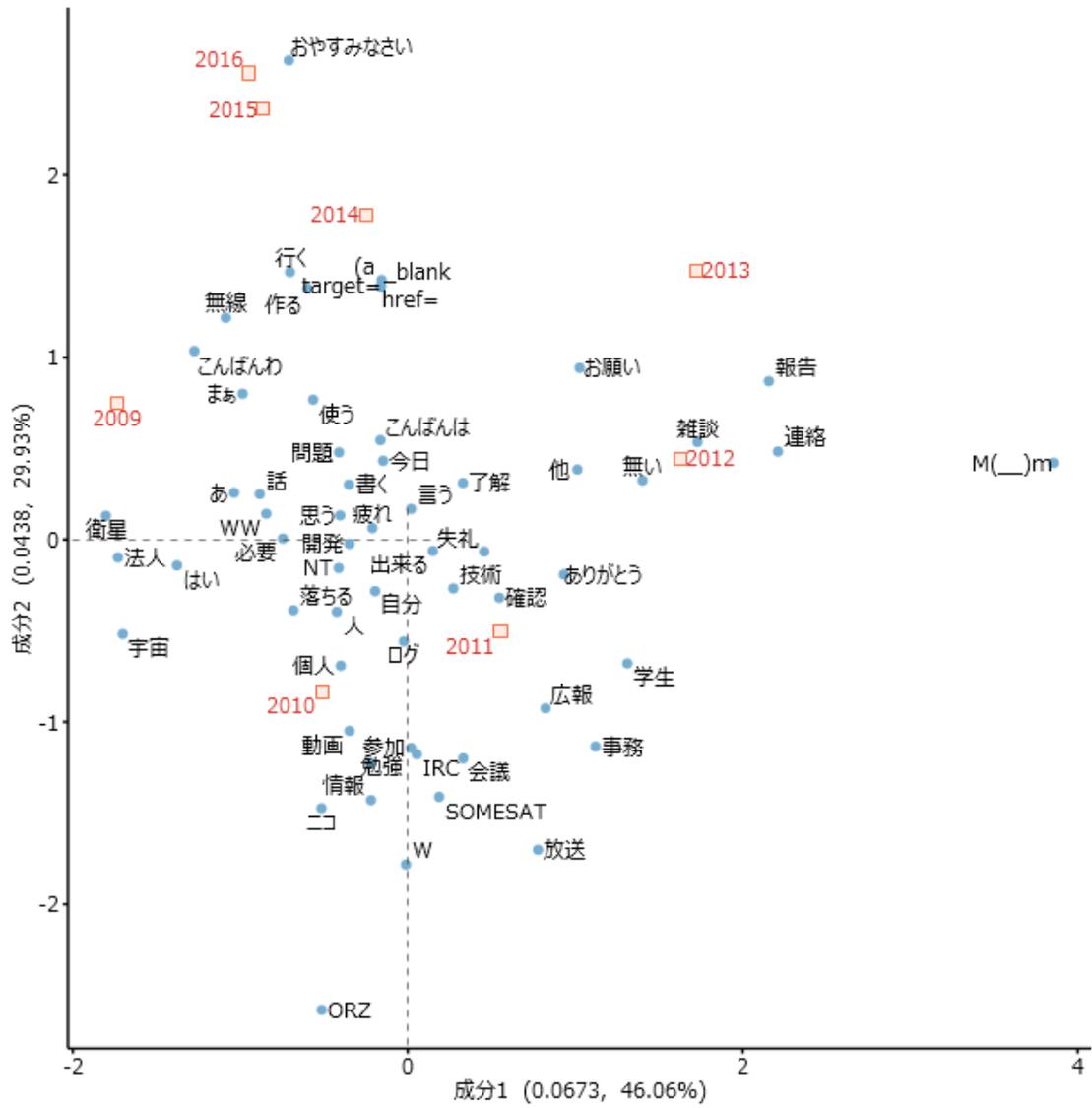


図 6. 対応分析 (筆者作成)

###背景### あなたは「テキストデータ分析の専門家」です。IRCのログ（テキストデータ）に対して、統計的な手法を用いて分析された結果を考察することができます。

###指令### こちらはSOMESAT/HAXAのIRCのログ（テキストデータ）をテキストマイニングした、対応分析という手法の分析結果です。赤い四角は変数（年）を表しています。以下の###対応分析の見方###を参考にし、この結果の解釈文を作成してください。

###対応分析の見方### 1) X軸の0から垂直方向に伸びる破線とY軸の0が水平方向に伸びる破線が交差しているところが原点です。2) 原点からある変数（赤い四角）の方向に向かって、その方向に原点から離れた語（青い丸）がその変数の特徴的な語と解釈されます。ただし、原点から変数（赤い四角）に向かって10°の角度の範囲に収まる語が特徴的な語であり、それ以外の語は特徴的な語ではないと判断します。3) 原点の近くに付置される語は、どの変数（赤い四角）にも偏らず、普遍的な語だと解釈されます。

図7. 対応分析の解釈用プロンプト（筆者作成）

年	代表語 (10°コーン内)	主に対応するテーマ	引用例
2009	衛星・法人・無線・ WWV・必要	テーマ6：衛星開発・ 宇宙技術 +法人格の議論	「HAXA 話は...大きな方向性は『衛星』と いうことで...」(2009-03-13) 「最悪 NPO 法人取りますか? >HAXA」 (2009-03-27)
2010	動画・勉強・参加・ 情報・ニコ	テーマ1：プロジェ クト参加と学習コミ ュニティ	「寒天培地動画作りたくてw」(2010-01- 30) 「理数系の勉強をしないですよー...」 (2010-01-30)
2011	技術・確認・出来 る・ログ・自分	テーマ4：ログ確認 /テーマ5：進捗報 告の芽生え	「これから定例報告会を行います。まずは 技術系から...」(2011-01-07) 「USB2.0 で...動作確認とれている...?」 (2011-01-04)
2012	雑談・連絡・他・ 無い	テーマ3：作業終わ り雑談+テーマ5： 連絡	「それでは、今日はここまで...後は雑談と したいと思います」(2012-01-07)
2013	報告・連絡・雑談	テーマ5：業務連絡・ 進捗報告が本格化	「技術系の皆様で相談・連絡・報告などお 願いします m(__)m」(2013-01-04)
2014	href=・target=・ _blank・行く・作 る	テーマ7：リンク貼 り付け/HTML タグ 断片	「」 (2014-01-03)
2015 -16	おやすみなさい	テーマ3：作業終わ りの挨拶が顕著(活 動量自体は減少)	「でわ、おやすみなさい。」(2015-01-03) 「おつかれさまでした&おやすみなさい」 (2016-01-02)

表 2. 年ごとの特徴語と7テーマの対応 (ChatGPT の応答を元に筆者作成)

4. まとめと示唆

1. 創成期（2009）は理想と技術（テーマ6）が先行し、まだ運営ルールは未整備。
2. 拡大・学習期（2010-11）で参加者が増え（テーマ1）、情報管理（テーマ2・4）への関心が芽生える。
3. 組織化期（2012-13）には定例会・報告書式（テーマ5）が確立し、雑談も枠組み化（テーマ3）。
4. ドキュメント化期（2014）に Web リンク共有（テーマ7）が急増し、非同期・リモート協働を下支え。
5. 終息・維持期（2015-16）は「おやすみなさい」中心の緩い雑談が主で、開発そのものは小休止。

対応分析の座標は、**左→右が「技術からマネジメントへ」、下→上が「作業中から後処理・終了へ」**という二軸を映し出しており、SOMESAT/HAXA IRC の8年間の

- 技術志向 → 組織運営志向
- 作業中心 → リンク共有 & 雑談中心
へ漸進した過程を定量的に裏付けている。

図8 対応分析の解釈に対する応答（一部）（ChatGPTの回答）

3. キーワード等の出現数の経年変化

(1) 書き込み数の経年変化

図9に書き込み数の経年変化を示す。2010年に58000回近いピーク値を記録した後は減少傾向にあることが分かる。

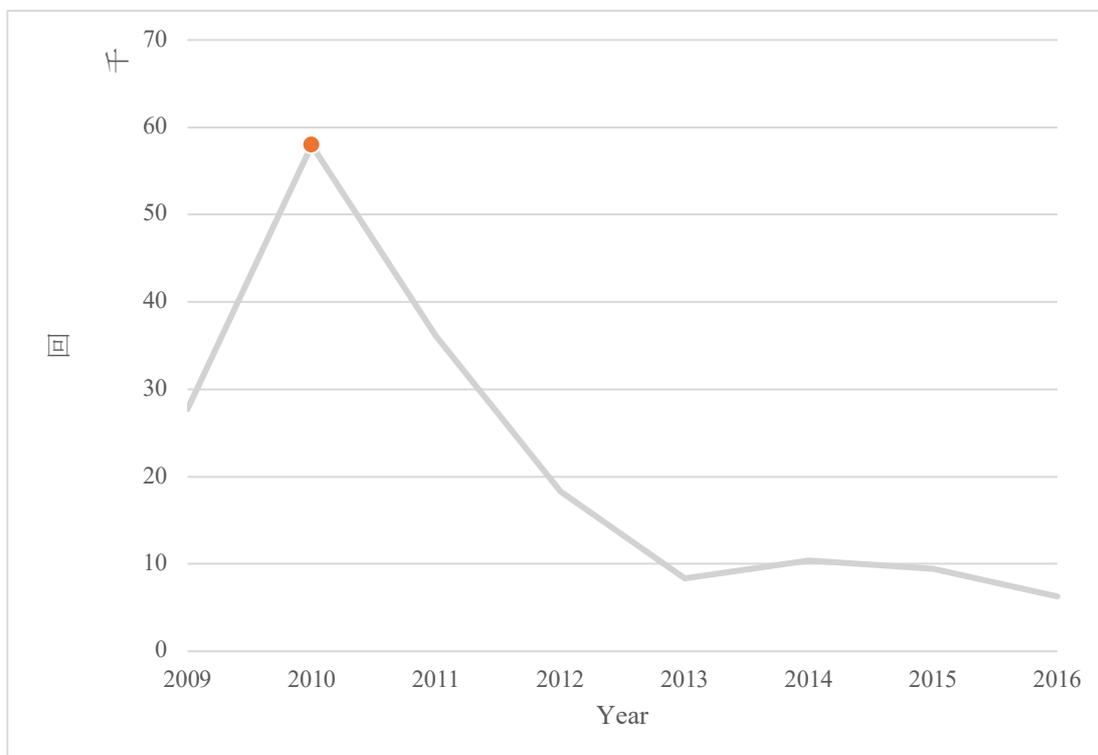


図9. 書き込み数の経年変化（筆者作成）

(2) ハンドルネーム毎の総書き込み数

図 10 にハンドルネーム（IRC に書き込んだ参加者名）毎の総書き込み数を示す。実際のハンドルネームはプライバシー保護のため伏せる。一部のハンドルネームに書き込み数が集中する、いわゆる「トールヘッド」型になっていることが分かる。

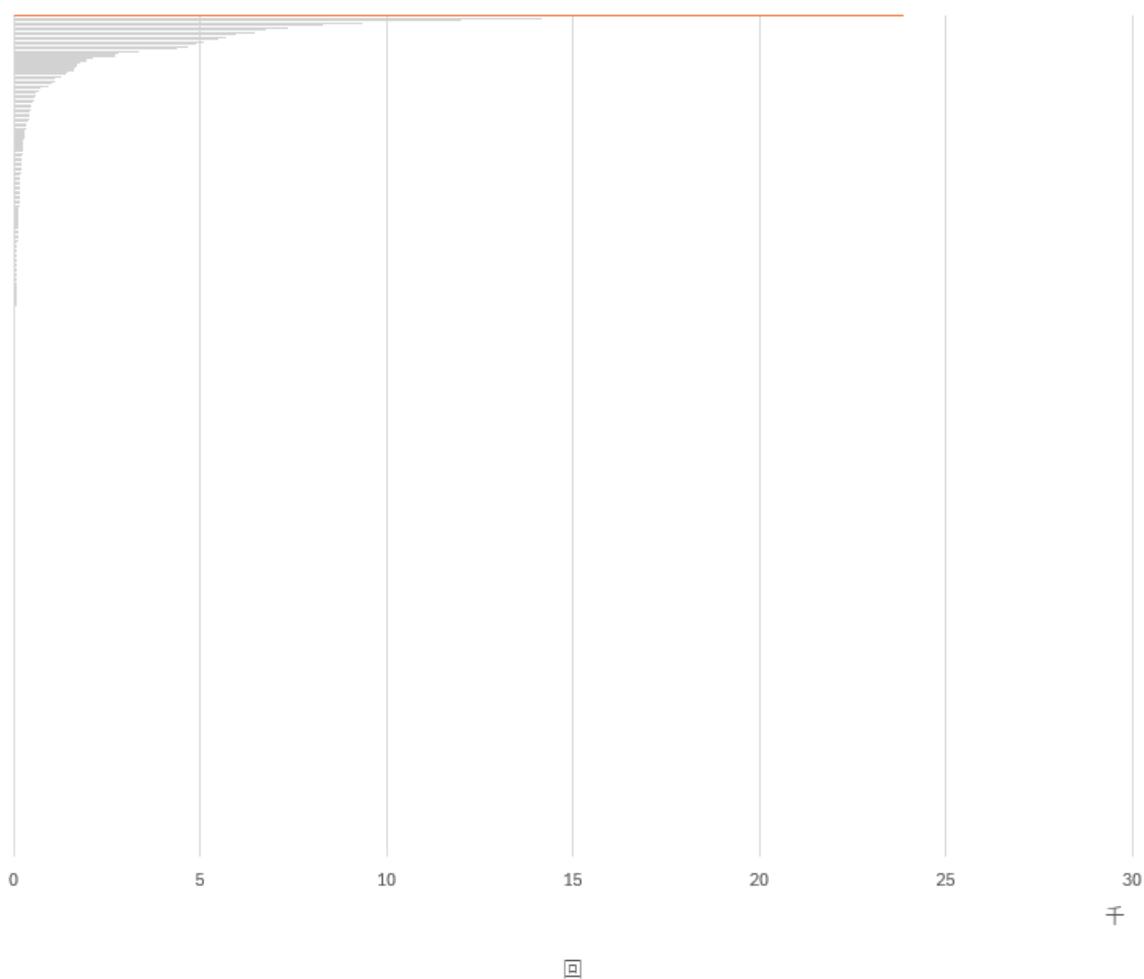


図 10. ハンドルネーム毎の総書き込み数（筆者作成）

(3) ハンドルネーム毎の書き込み数の経年変化

図 11 にハンドルネーム毎の書き込み数の経年変化を示す。実際のハンドルネームはプライバシー保護のため伏せる。色のついた各線が各参加者を表す。時間経過に伴って、議論や雑談を主導する参加者が交代している様子が見て取れる。

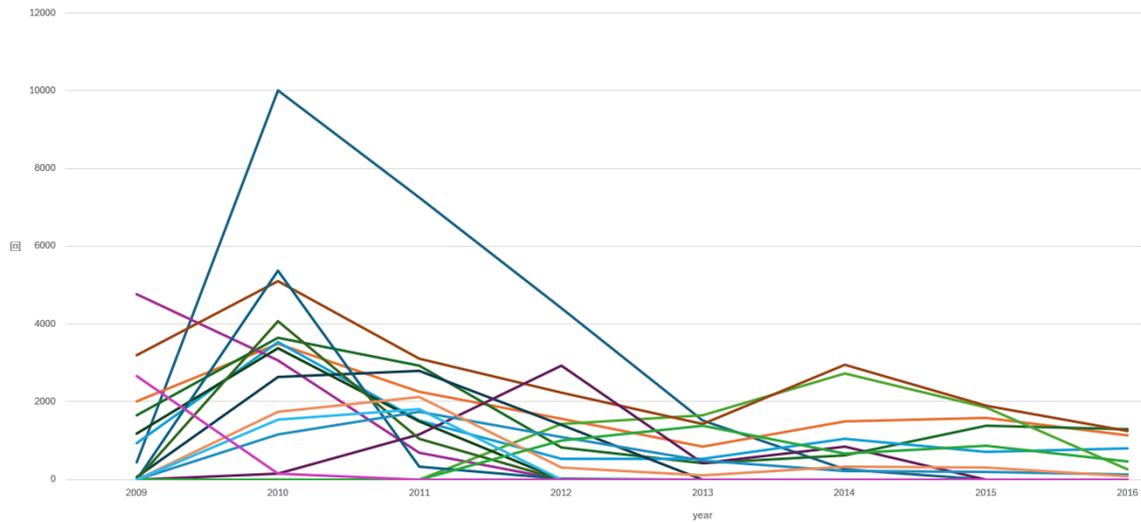


図 11. ハンドルネーム毎の書き込み数の経年変化 (筆者作成)

(4) 話題毎の書き込み数の経年変化

図 12 に話題毎の書き込み数の経年変化を示す。表記揺れを考慮し、幾つかの似通ったキーワードが出てくる話題を、同じテーマに関する話題としてカウントした。「初音」「はちゅね」「ミク」が出てくる話題は「初音等」に関する話題とした。「ネギ」「葱」が出てくる話題は「ネギ等」に関する話題とした。「衛星」「キューブサット」「cubesat」が出てくる話題は「cubesat 等」に関する話題とした。「カンサット」「缶サット」「cansat」が出てくる話題は「cansat 等」に関する話題とした。なお、「カンサット」とは、大学生等の宇宙人材の育成によく用いられている、模擬衛星のことである。「水ロケット」「水ロケ」「ペットボトルロケット」「モデルロケット」「高高度」が出てくる話題は「水ロケ等」に関する話題とした。「ZUNDA」「えだまめ」が出てくる話題は「ZUNDA 等」に関する話題とした。なお、「ZUNDA」や「えだまめ」は、SOMESAT がモデルロケットやカンサットの競技会に参加した際の、ロケットやカンサットに付けられた名称に関するキーワードである。図 12 からは、年を経るに従って関心事は変化していくものの、少なくとも 2016 年まではキューブサットが最大の関心事だったことが見て取れる。

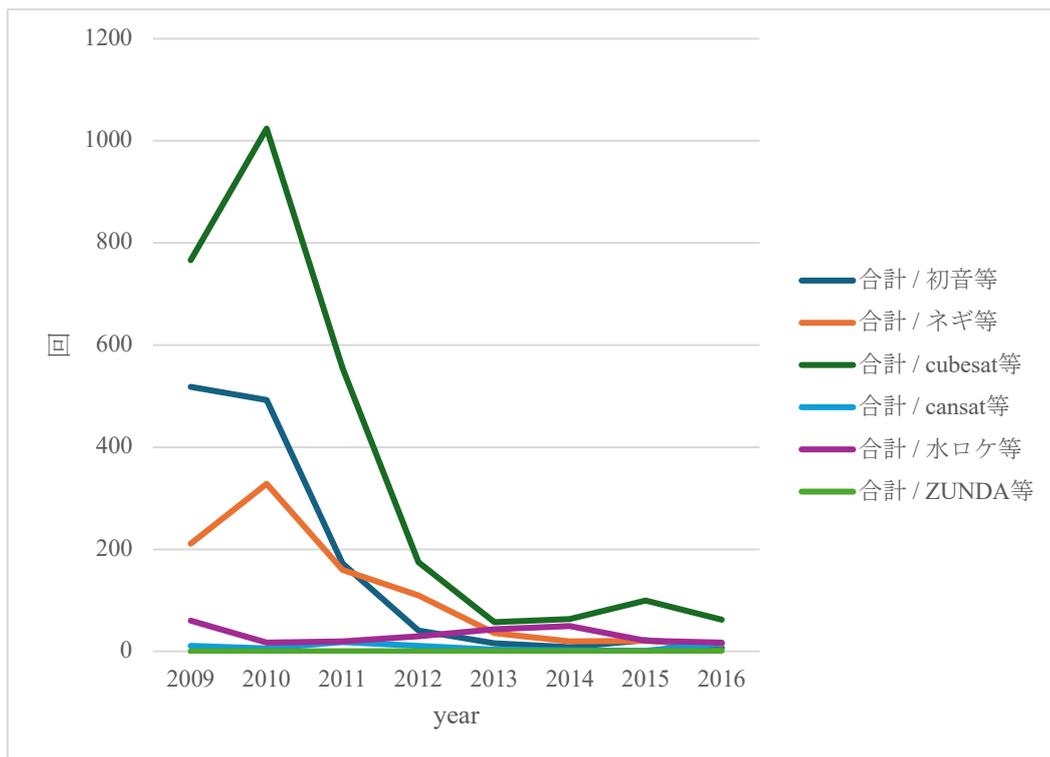


図 12. 話題毎の書き込み数の経年変化 (筆者作成)

IV. 考察とまとめ

解釈過剰で文脈に沿わなくなってしまう部分はあるものの、ChatGPT の解釈は概ね妥当であった。さらに言えば、ChatGPT による対応分析の解釈文 (図 8) にある「終息・維持期 (2015-16) は「おやすみなさい」中心の緩い雑談が主で、開発そのものは小休止」などといった記述

は、書き込み数の経年変化などや、本稿の範疇を超えるが著者が 2010 年から行っている SOMESAT の参与観察 (cf. 渡辺(2015), 渡辺・田邊(2016)) から裏付けられることであり、これが IRC のログの「年」を外部変数とした対応分析の結果のみから解釈できることは驚きである。だがこれは最初のプロンプトで詳細な文脈情報を与えたために可能になったことだと考えられる。詳細な文脈情報を LLM に与えるためには、やはり参与観察などの質的な研究方法が必要となる。ChatGPT の解釈は時に辻褃を合わせようとするあまり過剰なものになってしまうものの、KH Coder 等のテキストマイニングの再現可能な分析結果を人間が確認しつつ LLM の出力を一つのヒントとして活用すれば結果解釈の信頼性と質の底上げに寄与するという狙いは達成できたと言える。

謝辞

本研究にご協力くださったソーシャルメディア衛星開発プロジェクト SOMESAT の皆様には、心より感謝申し上げます。

文献

1. 周景龍, 中村康則, & 樋口耕一. (2024). KH Coder によって再現可能性を維持しつつ生成 AI から解釈のヒントを得る計量テキスト分析. 日本教育工学会研究報告集, 2024(4), 274–281.
2. 樋口耕一. (2020). *社会調査のための計量テキスト分析 [第2版] : 内容分析の継承と発展を目指して (第2版)*. ナカニシヤ出版.
3. 渡辺謙仁. (2015). 初音ミクと宇宙開発の草の根な関係 : 「ソーシャルメディア衛星開発プロジェクト SOMESAT」に着目して. 草の根文化の時代, 2, 33–53.
4. 渡辺謙仁, & 田邊鉄. (2016). 野火的な「プロジェクト」と学び : メディアとしての超小型衛星開発プロジェクトにおけるフィールドワークを通して. 認知科学, 23(3), 255–269. <https://doi.org/10.11225/jcss.23.255>

