一般発表 | Track 3: 情報検索・情報推薦・ソーシャルメディア

2026年3月2日(月) 13:00 ~ 15:10 | F会場

[8F] SNS分析(メディア/コンテンツ分析)

座長:劉 健全(日本電気)　コメンテータ:三林 亮太(神戸大学)

13:00 ~ 13:25

[8F-01] 音楽ストリーミングサービスの統計情報を用いた楽曲歌詞の季節性分析

*松村 歩空[1]、横山 昌平[1] (1. 東京都立大学)

13:25 ~ 13:50

[8F-02] Youtubeにおけるスポーツ応援スタイルの国・言語別分析

*高鍬 真輝[1]、横山 昌平[1] (1. 東京都立大学)

13:50 ~ 14:10

[8F-03] 動画コンテンツにおける話題区間の自動検出と人気度分析システムに関する研究

*小柳津 海来[1]、石橋 直樹[1] (1. 武蔵野大学)

14:10 ~ 14:30

[8F-04] Yahoo! 知恵袋に投稿された回答で参照されたwebページの役割の分析

*森高 楓[1]、松香 敏彦[1] (1. 千葉大学)

14:30 ~ 14:55

[8F-05] Steamユーザーの口コミを利用したゲームのプレイ体験の効果的な可視化

*相原 颯真[1]、横山 昌平[1] (1. 東京都立大学)

音楽ストリーミングサービスの統計情報を用いた楽曲歌詞の季節性分析

松村　歩空†　　横山　昌平†

† 東京都立大学　〒 191-0065 東京都日野市旭が丘 6-6
E-mail: † matsumura-hodaka@ed.tmu.ac.jp, shohei@tmu.ac.jp

あらまし　本研究は，音楽ストリーミングサービスの統計情報と楽曲の歌詞を用いて，歌詞に含まれる語彙の季節性が楽曲の流行時期とどのように関連しているのかを明らかにすることが目的である．そこで，音楽ストリーミングサービスの統計情報が実際に聴取された結果を反映するリアルタイムな動的指標の特性を最大限に活かし，従来の分析では見過ごされがちであった短期的な変動に着目する．研究アプローチとして，まず統計情報に時系列分析を適用し，楽曲の流行時期とその季節性を特定する．そして，この楽曲流行の季節性をもとに，歌詞から抽出した単語が示す季節的傾向を分析・検討することで，楽曲の流行と歌詞の季節的関連を解明し，現代の楽曲消費に現れる文化的な季節感を捉えることを目指す．

キーワード　時系列分析，STL 分析，統計情報，形態素解析

1　はじめに

日本で生活する多くの人々は，それぞれの季節に対して「春は出会いや期待」「夏は開放的で激しい」といった，ぼんやりとした共通のイメージを抱いている．こうした季節感は，単に気温や天候の変化だけで作られるものではない．卒業式や入学式，クリスマスといった，一年を通じて訪れる様々なイベントとも密接に結びついている．人々はこれらのイベントを経験する中で，特定の季節に対する情景や感情を強化し，共有している．音楽には，そのようなイベントや季節のイメージを鮮明に呼び起こすものがある．しかし，音楽と季節，あるいはイベントとの結びつきは，多くの場合，個人の感性や経験則といった主観的な文脈で語られるに留まってきた．人々の心にある「ぼんやりとした季節のイメージ」を客観的に捉えるためには，実際の楽曲消費においてそれがどのような言葉の重なりとして現れているのかを，実態に基づいた定量的な評価によって明らかにする必要がある．

楽曲の流行と季節の関連性を考える上で，楽曲が「いつ発売されたか」という固定的な情報だけでは，言葉が実際にいつリスナーに響いているのかという動的な実態を捉えることは困難である．クリスマスに過去の楽曲が再流行するように，リスナーの受容はリリース時とは限らず，日常生活の文脈に応じて変動するからである．現代の音楽視聴において主流となったストリーミングサービスの統計情報は，リスナーがその瞬間に楽曲を聴いている実績をリアルタイムに反映している．数十年前の楽曲であっても，特定のイベントや季節に合わせて繰り返しチャートへ浮上する現象などは，このデータによって数値として明確に捉えられるようになった．統計上の再生数の推移は，人々の意識がどの時期にどの楽曲へ向いているのかを示す直接的な指標といえる．

本研究は，このストリーミング統計から得られる「実際に楽曲が聴かれている時期」を，歌詞単語を分類するための物差しとして活用する．楽曲が流行している時期を統計的に特定し，それを時間軸上の基準とすることで，季節のイメージを実態に即して定量的に評価することが狙いである．

歌詞分析にあたっては，あらかじめ季節性が自明な単語に限定せず，歌詞に含まれるすべての単語を解析対象とする．これにより，伝統的な歳時記には載っていないような，現代の楽曲において特定の時期によく使われる意外な単語を発見できる可能性がある．国内の音楽チャートデータを活用した時系列分析と，それに基づく多角的な歌詞分析を通じて，現代の楽曲消費の中に立ち現れる季節感の解明を目指す．

2　関連研究

これまで，邦楽の歌詞が時代の変化や社会の様子をどう反映しているかを探る研究は数多く行われてきた．左古 [1] は，1968 年から 2013 年までの 46 年間にわたるヒット曲を分析し，歌詞に使われる言葉が時代とともにどう変化したかを調査した．また，長谷川ら [2] は，過去 60 年以上のヒット曲を対象に，その時代の出来事と歌詞の明るさに関連があるかを調べている．女性グループの歌詞に注目した小林ら [3] の研究や，特定のアーティストの語彙の変化を追った石川・川合 [4] の研究も，歌詞の計量的な分析手法を確立させてきた．これらの研究の多くは，オリコンの年間ランキングや CD の売上枚数といったデータをもとにしている．こうしたデータは「どの曲がどれだけ売れたか」を知るには適しているが，購入されたタイミングしか分からないという弱点がある．つまり，歌詞の研究において時代背景との関連は議論されてきたものの，リスナーがその曲を「いつ，どのようなタイミングで繰り返し聴いているか」という，日常の中での再生行動を詳しく捉えることは難しかった．そこで本研究では，音楽ストリーミングの統計情報を用いることで，リスナーが「いつ曲を再生したか」という詳細なデータを活用する．

音楽ストリーミングサービスの統計情報を用いた研究には，

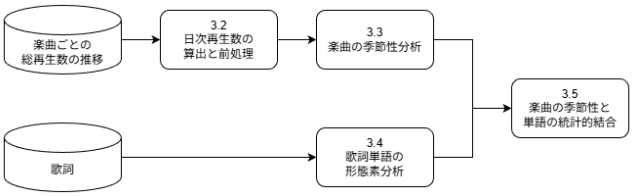


図 1: 分析の流れ

以下のようなものがある．Park ら [5] は，世界中の再生データを分析し，季節や時間帯によって好まれる「音の特徴（激しさや明るさ）」が周期的に変化することを突き止めた．日本国内でも，河野・石垣 [6] がストリーミングデータを使って各国の音楽の好みを比較している．また，Liew ら [7] は機械学習を用いてヒット曲の特徴を分類し，Machmudin ら [8] は時系列分析の手法を使って，楽曲の音響的なトレンドを分析している．これらのストリーミングデータを活用した最新の研究は，テンポや明るさといった「音の特徴」に注目したものがほとんどであり，歌詞の内容にまで踏み込んだものは極めて少ない．

佃・後藤ら [9] の調査によれば，スマホで音楽を聴く人の多くは，歌詞を深く理解したり共感したりするために，画面上で歌詞を頻繁に見ていることが分かっている．このことは，特定の季節に再生数が増える曲の裏側には，その季節にふさわしい「言葉」への需要があることを示唆している．本研究の独自性は，「音楽ストリーミングの統計情報」という行動データと，「歌詞」という意味データを掛け合わせた点にある．

3 分析手法

本章では，楽曲ごとの総再生数の推移データと歌詞データを用いて，季節特有の表現を抽出するための一連の処理プロセスについて述べる．

図 1 に分析の流れを示す．楽曲ごとの総再生数の推移データには，3.2 節で示す前処理と，3.3 節で示す季節性分析を行う．歌詞データには，3.4 節で示す形態素分析によって，単語に分離する処理を行う．得られた単語群と楽曲季節性を 3.5 節で述べる方法で結合することで，単語の季節性スコアを算出する．

3.1 データの収集と分析対象の選定

分析の母集団として，Spotify が公開している日次ランキング top200（Daily Top Songs Japan）に，2017/01/01 から 2025/09/30 までの間にランクインした実績を持つ楽曲を選定した．しかし，ランキングデータのみではランク圏外期間の推移が不明である．そこで，外部の音楽データプラットフォーム Chartmetric を利用して，各楽曲の総再生数の推移を取得した．

取得した 8,000 曲を超える楽曲に対し，本研究の目的である日本における歌詞の季節性分析に適したデータセットを構築するため，データ処理に先立ち以下の基準でフィルタリングを実施した．

まず，アーティスト名・楽曲名・歌詞のいずれかに日本語を含むのみを抽出した．これにより，歌詞のない楽曲や，歌詞が完全に外国語である楽曲を除外した．

次に，派生楽曲を除外した．同一楽曲の「Remix 盤」「Acoustic Ver.」「Live Ver.」や「カバー楽曲」等の派生楽曲が含まれる場合，同一の歌詞がデータセット内に重複して存在することになる．特定の単語頻度が不当に高く評価される統計的バイアスが生じるため，これらの派生楽曲を除外し，オリジナル音源のみを分析対象として選定した．オリジナル音源がランクインしていない場合は，派生楽曲を一つ採用した．

これらのフィルタリングにより，対象楽曲は約 3,000 曲となった．

また，歌詞に含まれる単語を分析するにあたり，基盤となる歌詞データを収集した．歌詞データについては，公式に発表されている情報を最優先とし，複数の歌詞検索サービスや配信プラットフォームの情報を補完的に用いることで，信頼性と網羅性を確保した．

3.2 日次再生数の算出と前処理

3.1 節で選定した楽曲群について，総再生数データから季節変動を観測するための日次時系列データへの変換を行った．Chartmetric の時系列データには，プラットフォーム側の集計遅延や通信エラーに起因する観測ノイズが含まれており，そのままでは季節性やトレンドの解析が困難であった．具体的には，再生数が記録されない不定期な欠損，その欠損期間の再生数が翌日以降にまとめて計上されることで発生する，本来の推移とは不整合な一時的突出値（スパイク），および記録日が数日前後するタイミングのズレである．これらのノイズに対し，単純な移動平均や外れ値の除外処理を行うことは，楽曲の総再生数の整合性を損なうだけでなく，バイラルヒットなどによる急激なトレンドの変化をも誤ってノイズとして処理してしまうリスクがある．

したがって本研究では，Chartmetric から取得した総再生数の推移に対して，総再生数の整合性を維持しつつ，本来の日次変動を復元する独自の多段階再構築アルゴリズムを構築し，適用した．

3.2.1 異常検知とスキップ探索

第一段階として，各データ点に対する異常検知を行った．ここでは，連続する異常値による誤検知を防ぐための比較基準として，各データ点に対し前後 7 日間の各中央値を「ローカルベースライン」として算出した．

また，異常値の連続による誤判定を防ぐため，「スキップ探索」を導入した．これは，比較対象となる隣接データ自体が欠損やスパイク候補である場合にはそれを参照せず，さらに近傍にある正常な値を探索して比較を行う手法である．

異常値の判定基準としては，統計的な外れ値検出で一般的に用いられる Tukey の基準（四分位範囲の 1.5 倍則）に準拠し，比較対象の 1.5 倍以上の変動をスパイク候補とした．さらに，スパイク判定閾値（1.5 倍）の逆数（$1/1.5 \approx 0.67$）を基準とした 0.7 倍を下限値として設定し，補正後の値が比較対象の 0.7 倍を下回る場合は，分割を棄却した．これにより，本来のトレンドによる急上昇と，システムエラーによるスパイクを区別している．

3.2.2 制約付き時系列シフトによる全体最適化

第二段階では，検出された欠損とスパイクに対し，それらが記録タイミングのズレによって生じた同一の事象であると仮定し，時系列のシフトによる修復を行った．欠損とスパイクのペアを探索し，その間のデータを数値を変えずに日付インデックスのみ移動させることで，スパイクとして過剰計上された分を物理的に欠損日へ還元した．

この処理においては，全体最適化のアプローチを取り入れている．各データ点の累積移動量が元のタイムスタンプから前後5日を超えないという制約条件下で，最も距離の近いペアから順に処理を実行した．この手法により，実測データの波形情報，すなわちユーザーの聴取行動を破壊することなく，データを本来あるべき日付へと近づけた．

シフト処理で解決しきれなかった残存する欠損区間に対しては，PCHIP（Piecewise Cubic Hermite Interpolating Polynomial）法を用いて，前後のトレンドに滑らかに接続するように補完した．

図2に，前処理前後の日次再生数推移例を示す．図2(a)はクリスマスがテーマの楽曲A，図2(b)は上位にランクインし続けた楽曲Bの例であり，各グラフの灰色の線が前処理前，青色の線が前処理後のデータである．前処理によって観測ノイズが除去され，楽曲の実際の聴取傾向をより正確に反映した時系列データが得られた．

3.3 楽曲の季節性分析

抽出された楽曲の日次再生数推移には，「季節による周期的な再燃（季節性）」だけでなく，「リリース直後の一過性の流行（バズ）」や「長期的な人気の上昇・下降（トレンド）」が混在している．これらを数学的に分離し，真に季節性を持つ楽曲のみを特定するために，以下の手順で時系列分析を行い，季節性を確認した．

3.3.1 月次集約とリリース初期のスパイク抑制

まず，季節変動の長期的な傾向を捉えるため，日次データを月次平均に集約した．

次に，リリース直後のプロモーションや話題性による爆発的な再生数増加（スパイク）の処理を行った．この非周期的な急騰をそのまま分析すると，トレンド推定が歪む原因となる．そのため，データ開始後6か月間の最大値が，その後の安定期（6か月目〜12か月目）の中央値の3倍を超える場合，初期の急騰を一過性のものと判定した．該当楽曲については，開始後6か月間のデータを安定期の中央値で置換する平滑化処理を行った．

3.3.2 対数変換とSTL分解

再生数は楽曲によって規模が大きく異なり，分散不均一性を持つため，全てのデータに対数変換を施した．時点 t における日次平均再生数を x_t とすると，変換後の値 y_t は式(1)で表される．

$$y_t = \log(x_t + 1) \tag{1}$$

分解手法には，外れ値に対して堅牢なSTL分解（Seasonal-Trend decomposition using LoESS）を採用した．本研究では式(2)で表される加法モデルを仮定し，時系列データ y_t をトレンド成分 T_t，季節成分 S_t，残差成分 R_t に分解した．

$$y_t = T_t + S_t + R_t \tag{2}$$

パラメータ設定と根拠は以下のとおりである．

- 周期 (Period) = 12: 月次データの年周期性を捉えるため．
- 季節平滑化窓 = 7: 本データの観測期間において，単年の特異な変動に過学習せず，かつ経年による緩やかな変化を許容する値として設定．
- トレンド平滑化窓 = 13: 12か月の周期変動を完全に相殺し，トレンドへの季節成分の混入を防ぐ最小の奇数として設定．

これにより，短期的なノイズの影響を受けずに，年単位の緩やかな季節変化 S_t を抽出した．

図3に，図2で示した楽曲Aおよび楽曲BのSTL分解結果を示す．

3.3.3 トレンド除去後の複数年検証

抽出された季節成分 S_t が，単なる偶然や一回きりの事象でないことを確認するため，トレンドを除去した実データを用いた再現性検証を行った．まず，季節成分の振幅を統一的な基準で評価するため，全期間における季節成分の平均 μ_S と標準偏差 σ_S を用いて標準化（Zスコア化）を行った．時点 t における標準化スコア Z_t は式(3)で定義される．

$$Z_t = \frac{S_t - \mu_S}{\sigma_S} \tag{3}$$

この Z_t が閾値0.8を超える時点が，複数年にわたって観測されるかを確認した．具体的には，$Z_t > 0.8$ を満たす月が，翌年以降も同じ時期（前後1か月を含む3か月間）に観測されるかを調査した．再燃が観測された楽曲のみを「再現性のある季節性楽曲」と判定し，それ以外の楽曲は除外した．0.8は，正規分布の上位約20%に相当し，1年のうち約2.5ヶ月〜3ヶ月の期間でピークを形成する波形を抽出するための設定である．

3.3.4 観測期間バイアスの補正

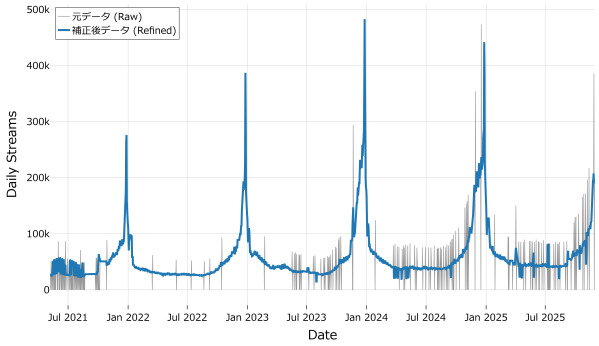
最終的に得られた季節性成分に対し，月ごとの代表値を算出した．ある月 m（$m = 1, \ldots, 12$）に属する時点集合を $\mathcal{T}_m$ とし，その要素数を N_m とするとき，月次平均スコア $\bar{Z}_m$ は式(4)のように求められる．

$$\bar{Z}_m = \frac{1}{N_m} \sum_{t \in \mathcal{T}_m} Z_t \tag{4}$$

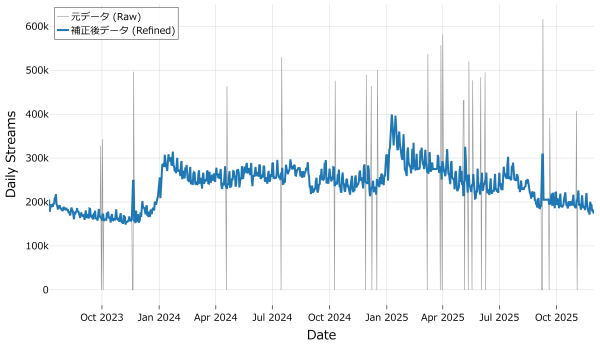
ここで，データの観測期間が12か月の整数倍でない場合，特定の月のデータが多く含まれることで，全月の総和が0にならないバイアスが発生する．この問題を解消するため，月ごとのスコアから全月の平均値を減算する処理を行った．最終的な補正済み季節性スコア Z'_m は式(5)で与えられる．

$$Z'_m = \bar{Z}_m - \frac{1}{12} \sum_{k=1}^{12} \bar{Z}_k \tag{5}$$

この補正により，$\sum_{m=1}^{12} Z'_m = 0$ が成立し，観測期間の長さに関わらず，公平な季節性評価を可能とした．

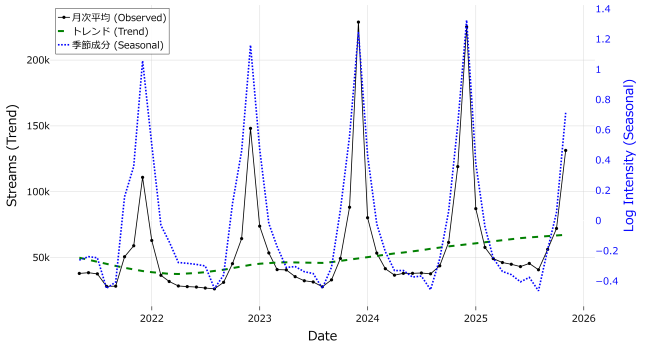


(a) 楽曲 A

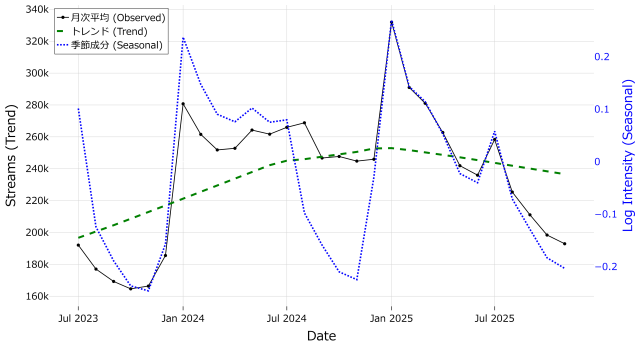


(b) 楽曲 B

図 2: 前処理後の日次再生数推移例



(a) 楽曲 A



(b) 楽曲 B

図 3: STL 分解の結果

3.4 歌詞単語の形態素分析

歌詞テキストに対し，形態素解析エンジン SudachiPy を用いた単語分割を行った．本研究では，単一の語彙だけでなく「夏休み」のような複合名詞を適切に一つの単位として抽出するため，分割強度の最も高い「Mode C」を採用した．

さらに，解析結果を本研究の目的に適応させるため，独自の解析辞書による制御ルールを適用した．特定の固有名詞や連語が不自然に分割されるのを防ぐ「MERGE」処理，表記揺れを代表語に統合する「REPLACE」処理，および分析上のノイズとなる音符記号や記号類，頻出する代名詞や自立語などを除外する「STOP」処理の 3 段階である．抽出対象とする品詞は，意味内容を強く保持する名詞，動詞，形容詞，形状詞に限定した．

3.5 楽曲の季節性と歌詞単語の統計的結合

3.3 節で得られた楽曲単位の季節性スコアと，3.4 節で抽出された歌詞単語の情報を数学的に結合し，単語単位での季節性指標を確立する手法について述べる．

3.5.1 楽曲から単語への季節性スコアの分配

楽曲の季節的な流行特性を単語へ継承させるため，楽曲が持つ月別の季節性スコアを，その楽曲を構成する単語群へと分配した．ある楽曲 k に含まれる単語 w の月 m における分配スコア $S_{w,k,m}$ は，楽曲の季節性スコア $S_{k,m}$ および楽曲内での単語出現回数 $C_{w,k}$ を用いて式 (6) のように定義される．

$$S_{w,k,m} = \begin{cases} S_{k,m} \times (1 + \log_2 C_{w,k}) & (S_{k,m} > 0) \\ 0 & (S_{k,m} \leqq 0) \end{cases} \tag{6}$$

出現回数に対して対数重みを適用することで，楽曲内での連呼によるスコアの偏りを抑制した．

また，楽曲スコアが正の値を示す月，すなわち「その楽曲が統計的に季節性を帯びている月」のみを分配の対象とした．単語 w の最終的な月別スコア $S_{w,m}$ は，その単語が出現する全楽曲においてこれらの値を合算することで得られる．

3.5.2 単語単位の季節性評価指標の定義

算出された月別スコアに基づき，単語の季節的特性を測る指標を定義した．単純なスコアの振幅による評価では，季節性に関わらず出現頻度が極めて高い単語が，微小な変動であっても大きなスコアを得てしまう．そこで，式 (7) のように単語 w の各月のスコア $S_{w,m}$ をその単語の全月平均スコア $\bar{S}_w$ で除した「リフト値（特化係数）」$L_{w,m}$ を導入した．

$$L_{w,m} = \frac{S_{w,m}}{\frac{1}{12}\sum_{i=1}^{12} S_{w,i}} \tag{7}$$

$$\tag{8}$$

この定義によれば，年間を通じて一定の頻度で出現する単語は $L_{w,m}$ が常に 1.0 付近に収束する．一方，特定の月に集中して出現する単語は，その月において $L_{w,m}$ が大きくなり，他の月では 1.0 を大きく下回る値を示す．このリフト値を用いるこ

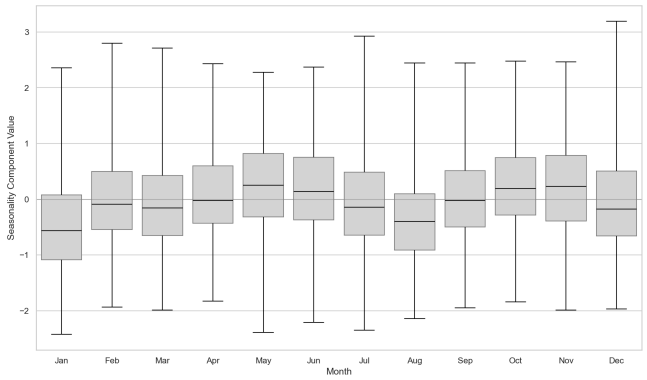


図 4: 季節性成分の月別分布と四分位数

表 1: 各月における季節性スコアの分布の内訳

月	$Z \leqq -3$	$Z \leqq -2$	$Z \leqq -1$	$Z \leqq 0$	$Z \geqq 0$	$Z \geqq 1$	$Z \geqq 2$	$Z \geqq 3$
1 月	0	16	508	1285	492	58	7	0
2 月	0	0	103	971	806	125	12	0
3 月	0	0	171	1039	738	136	14	0
4 月	0	0	53	907	870	228	10	0
5 月	0	37	126	665	1112	353	13	0
6 月	0	2	125	765	1012	302	15	0
7 月	0	4	147	1001	776	174	12	0
8 月	0	3	371	1272	505	121	6	0
9 月	0	0	117	926	851	217	12	0
10 月	0	0	68	694	1083	301	19	0
11 月	0	0	116	741	1036	328	29	0
12 月	0	0	202	1020	757	198	58	7

とで，季節性に特化した単語を公平に評価できる．

4 分析結果

本章では，前章で述べた分析手法によって得られた楽曲及び単語の季節性スコアについて，統計的な結果とそれに対する考察について記述する．

4.1 楽曲の季節性

3.3 節 で行った楽曲の季節性分析によって，楽曲ごとに，各月の季節性スコアが算出できた．図 4 は，箱ひげ図を用いて補正済み季節性スコア Z'_m の分布を示したものである．また，表 1 に，Z'_m が特定の閾値を超えた楽曲数を月ごとに集計した結果を示す．季節性の強さを段階的に評価するため，閾値は正の方向（$0, \geqq 1, \geqq 2, \geqq 3$）および負の方向（$0, \leqq -1, \leqq -2, \leqq -3$）のそれぞれについて確認した．表内では，簡略化のために，Z'_m を Z として記載している．

図 4 および表 1 の集計結果からは，楽曲の再生傾向がランダムではなく，5 月ごろと 11 月ごろに二つの山を持つ明確な年間周期を有していることが確認できた．さらに，季節性成分の正負（0 以上/0 以下）という全体的な傾向を含めて分析することで，各月の特性がより詳細に明らかとなった．

年間の流れを追うと，まず 1 月の大きな落ち込みを底として，3 月から 5 月にかけて正の季節性を示す楽曲数が増加する第一の上昇局面が見られる．特に 5 月は，閾値 1 以上を示す強い季節性の楽曲数が 353 曲と年間最多であるだけでなく，$Z \geqq 0$（正の季節性を持つ楽曲数）においても 1112 曲と全月の中で最大値を記録している．これは，5 月が特定の少数の曲に限らず，分析対象とした楽曲の多くが広く聴かれやすい，最も「広範な上昇トレンド」にある時期であることを裏付けている．ただし，$Z \leqq -2$ という強い季節性の減少を示す楽曲も一部存在しており，全体としては好調ながらも楽曲ごとの二極化が生じやすい時期でもある．

その後，夏期にかけては全体的な停滞が見られる．特に 8 月は，$Z \leqq -1$ の楽曲数（371 曲）が多いことに加え，$Z \leqq 0$ の楽曲数が 1272 曲に達しており，これは 1 月に次ぐ年間 2 位の多さである．「夏ソング」と呼ばれる特定のジャンルが存在する一方で，大多数の楽曲にとっては再生数が伸び悩む時期であることが，全体数のデータからも示唆された．

秋季からは再び上昇局面に転じ，10 月から 11 月にかけては $Z \geqq 0$ の楽曲数が再び 1000 曲を超える「第二の広範な上昇期」を迎える．しかし，続く 12 月は様相が一変する．12 月は $Z \geqq 2$ や $Z \geqq 3$ といった極めて高い値を示す楽曲が集中しており，年間で最も強い季節性バーストが観測される．その一方で，$Z \leqq 0$ の楽曲数は 1020 曲に達し，$Z \geqq 0$ の 757 曲を大きく上回っている．すなわち 12 月は，クリスマスソング等の特定楽曲が爆発的に再生される一方で，それ以外の一般的な楽曲の多くは再生機会を奪われ，減少傾向（負の季節性）に転じるという，「極端な集中と選別」が発生している特異な月であると言える．

そして年が明けた 1 月には，$Z \leqq 0$ の楽曲数が 1285 曲（全月最多），$Z \leqq -1$ の楽曲数も 508 曲（全月最多）となり，あらゆる指標において顕著な減少が観測された．これは 1 月が，ごく一部の例外を除き，年間を通じて最も楽曲の再生回数が伸びにくい時期であることを示している．

以上の結果より，本データにおける季節性は一様ではなく，春・秋に見られる「多くの楽曲が緩やかに盛り上がるフェーズ」と，夏・冬に見られる「一部の楽曲のみが突出する（あるいは全体が沈静化する）フェーズ」という，質的に異なる統計的特性が循環していることが明らかとなった．

4.2 単語の季節性

3.4 節で行った楽曲の季節性と歌詞単語の統計的結合によって，単語ごとに各月の季節性スコアが算出できた．本分析からは，季節に対する人々の「ぼんやりとしたイメージ」が，実際には具体的なイベントや心理状態と密接に結びついていることが明らかとなった．表 2 は，結果に特徴が表れていた月の，リフト値の高い特徴語トップ 5 を示したものである．各単語の下には，リフト値を有効数字三桁で記載している．

第一に確認されたのは，季節のイメージと伝統的なイベントとの結びつきである．表 2 に示す通り，12 月においては「クリスマス」のリフト値が 6.15 と突出して高く，「雪」「ケーキ」といった関連語が上位を占めている．これは，「冬＝クリスマス」というイメージが実際の楽曲消費行動において圧倒的な支配力を持っていることを統計的に裏付けるものである．同様に，3

表 2: 歌詞における月別特徴語とリフト値

順位	3 月	4 月	5 月	8 月	11 月	12 月
1	春風	桜	現在	渚	クリスマス	クリスマス
	(3.37)	(3.54)	(3.43)	(3.22)	(4.01)	(6.15)
2	旅立つ	春風	何色	向日葵	近づく	雪
	(2.72)	(3.34)	(3.38)	(2.32)	(2.83)	(4.32)
3	桜	革命	挑む	花火	プレゼント	降り積もる
	(2.48)	(2.61)	(3.38)	(2.32)	(2.77)	(4.32)
4	母	花弁	諦め	DJ	冷静	ケーキ
	(2.43)	(2.54)	(3.37)	(2.18)	(2.72)	(3.70)
5	仰ぐ	母	主役	曲	寒い	今年
	(2.29)	(2.41)	(3.30)	(2.13)	(2.55)	(3.21)

月・4 月における「春風」「桜」といった単語の抽出は，四季の変化や花見といったイベントが，リスナーの再生行動と強く同期していることを示している．

第二に，伝統的な歳時記には見られない季節語や感情の発見である．3 月のランキングにおける「旅立つ」「母」の出現は，単なる春の訪れ以上に，進学や就職に伴う「家族との別れと感謝」がこの季節の主要な感情テーマであることを示している．また，8 月における「dj」「曲」の出現は，音楽フェスなどのイベント文化が現代の夏の風物詩として定着していることを示唆する．さらに特筆すべきは 5 月であり，気候を表す言葉の代わりに「挑む」「諦め」といった葛藤を表す語彙が上位を占めた．これは，五月病のような新生活の緊張が解けた後に訪れる心理的な揺らぎに寄り添う楽曲が，この時期に強く求められているという，現代社会特有の季節感を表していると解釈できる．

また，全体的な傾向として，季節によって「強さ」の質が異なることも判明した．12 月は「クリスマス」等の特定単語が極めて高いリフト値を示す「一点突破型」の強さを持つのに対し，5 月は心理描写に関連する多数の単語が高い数値を維持する「面での広がり」を持つ強さが見られた．一方，夏（8 月）は文化的イメージの強さに反してリフト値のピークは相対的に低く，歌詞における季節性は冬や春ほど先鋭化していないことが示唆された．

5 まとめと今後の課題

本研究では，ストリーミングサービスの再生履歴データと歌詞情報を統合的に解析することで，現代の楽曲消費における季節性の実態を定量的に解明することを試みた．まず，STL 分解を用いた時系列解析により，楽曲の再生傾向には春季と冬季にピークを持つ明確な年間周期が存在することを確認した．特に 12 月は一部の楽曲への極端な集中が見られる「局所的ピーク」であり，5 月は広範な楽曲が聴かれる「全体的トレンド」であるというように，時期によって季節性の性質が異なることが明らかとなった．

次に，この統計的な季節変動を基準として歌詞の単語分析を行った結果，12 月の「クリスマス」や 4 月の「桜」といった伝統的な季節語が，実際の再生行動と極めて強く同期していることが裏付けられた．さらに，5 月における「挑む」「諦め」や，8 月における「dj」といった，従来の歳時記にはない現代特有の季節語を発見することに成功した．これらは，新生活後の心理的葛藤やフェス文化といった現代の文脈が，リスナーの楽曲選択に深く影響していることを示唆している．以上の結果より，ストリーミングデータと歌詞分析を組み合わせる本手法は，これまで主観的に語られがちであった音楽と季節の結びつきを，客観的かつ定量的な指標として提示することに成功したと言える．

一方で，本研究にはいくつかの課題が残る．

一つ目は，「季節性」の定義に関する問題である．本研究では時系列解析によって抽出された周期的なヒットを季節性と定義したが，この手法ではアーティストの記念日やイベント露出など，季節とは直接関係のない要因による周期的なヒットを誤検出する可能性がある．

二つ目は，歌詞分析の解像度である．単語単位の頻度分析では，比喩表現や否定語による意味の反転を区別できていない．より精緻な解析のためには，係り受け解析や高度な自然言語処理モデルを用いた意味解析が必要となる．

今後の展望として，気象データや SNS トレンド，アーティストの活動履歴といった外部データをモデルに統合することが挙げられる．多角的な要因を考慮した分析を行うことで，単なる数値上の周期性を超え，音楽が人々の生活や感情に寄り添うメカニズムをより本質的に解明できると期待される．

文　献

[1] 左古輝人. ヒットソング歌詞の変遷：1968 年から 2013 年まで. 人文学報, No. 497, pp. 49–85, 2015.
[2] 長谷川大和, 岡田龍太郎, 田丸翔大, 朝倉大樹. ヒット曲の歌詞データの感情分析による年別感情推移の可視化方式および感情と出来事の関係性の調査. 情報処理学会第 86 回全国大会講演論文集, 2024.
[3] 小林佳織, 狩野恵里奈, 鈴木崇史. 女性グループの歌詞の計量テキスト分析. 言語処理学会第 19 回年次大会発表論文集, 2013.
[4] 石川琴美, 川合康央. 日本語楽曲を対象とした歌詞の時系列変化分析. 情報システム学会第 16 回全国大会研究発表大会論文集, 2020.
[5] Minsu Park, Jennifer Thom, Sarah Mennicken, Henriette Cramer, and Michael Macy. Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, Vol. 3, No. 3, pp. 230–236, 2019.
[6] 河野瑞樹, 石垣司. 音楽ストリーミングサービスで配信された楽曲の特徴量を用いた音楽嗜好性の国際比較. サービソロジー論文誌, Vol. 7, No. 1, 2023.
[7] Kongmeng Liew, Yukiko Uchida, and Igor de Almeida. Cultural differences in music features across taiwanese, japanese and american markets. *PeerJ Computer Science*, Vol. 7, p. e642, 2021.
[8] Daffa Ghifari Machmudin, Mila Novita, and Gani Ardaneswari. Analysis of spotify's audio features trends using time series decomposition and vector autoregressive (var) model. In *International Conference on Data Science and Official Statistics (ICDSOS)*, 2023.
[9] Kosetsu Tsukuda, Masahiro Hamasaki, and Masataka Goto. Why and how people view lyrics while listening to music on a smartphone. *IEICE Transactions on Information and Systems*, Vol. E106-D, No. 4, 2023.

Youtubeにおけるスポーツ応援スタイルの国・言語別分析

高鍬　真輝†　　横山　昌平††

† 東京都立大学システムデザイン学部　〒 191–0065 東京都日野市旭が丘 6-6
†† 東京都立大学大学院システムデザイン研究科　〒 191–0065 東京都日野市旭が丘 6-6
E-mail: †takakwa-masaki@ed.tmu.ac.jp, ††shohei@tmu.ac.jp

あらまし　近年，スポーツ観戦は多様化しており，特に SNS を通じたリアルタイム観戦が国内外で普及している．これに伴い，SNS のデータからサッカーの試合の観戦スタイル，ファンの感情変化を分析をしている研究が増加している．しかし，既存研究の多くは，単一言語圏内の分析のみにとどまっており，国際的な観戦スタイルの比較や文化的差異の検証については十分行われていない．そこで，本研究では YouTube のライブ配信から取得したコメントデータを用いて，BERT 感情分析モデルおよび BERTopic トピックモデリングを適用することで，試合中の感情変化の推移について時系列で可視化し，言語・地域別の観点から観戦スタイルの差異を明らかにした．

キーワード　SNS，スポーツ，コメントデータ，BERT

1　はじめに

近年，SNS の発展により配信プラットフォームを通したスポーツの観戦が人気を博している．笹川スポーツ財団によると，スタジアムでの現地観戦，テレビによるスポーツ観戦率は年々下がっているものの，インターネットによるスポーツ観戦率は 2018 年の 11.5%から 2024 年には 24.2%まで上昇している [1]．特に，「YouTube」や「X」といった SNS を通じたリアルタイム観戦が急速に普及している．これらは単に映像を見るだけでなく，コメント機能を通じて他の視聴者と感情を共有したり，実況配信者と双方向のコミュニケーションを取ったりできる点に特徴がある．物理的に離れた場所にいるファン同士が，あたかも一つのスタジアムにいるかのような一体感（バーチャルな熱狂）を生み出すことができるため，コメントデータには視聴者の生の感情が色濃く反映されている．また，スポーツの応援スタイルや楽しみ方は，国や地域ごとの文化背景によって大きく異なることが知られている．顕著な具体例として野球が挙げられる．日本のプロ野球では，選手ごとに専用の応援歌が作られ，応援団の主導に合わせて観客全員で合唱やコールを行う組織的な応援が一般的である．対照的に，アメリカのメジャーリーグ（MLB）ではそのような集団的な応援文化はなく，観客が個々に拍手を送ったり，あるいは野次を飛ばしたりと，個人の楽しみ方が尊重される傾向にある．このような国・地域ごとの観戦態度の違いは，当然ながら他競技や SNS 上のコメント行動にも反映されると考えられる．これに伴い，SNS のコメントデータを用いた観戦者の感情の推移を分析した研究が行われている．

例えば，試合中のコメントからファンのネガティブ・ポジティブといった感情を抽出する試みや，トピックモデルを用いて主要な話題について分析する研究が行われている．しかしながら，既存の多くの研究は単一言語圏を対象としており，言語・地域別から見た観戦態度の違いの検証は十分ではない．さらに，試合全体の分析のみにとどまるものが多く，試合展開に連動した詳細な時系列での感情変化までは明らかにされていないことが多い．

そこで，本研究では 2025 年 10 月 14 日に行われたサッカーキリンチャレンジカップのブラジル対日本の YouTube ライブ配信コメントを英語，ポルトガル語，日本語の 3 言語から取得した．データソースとして，それぞれ異なる応援文化や背景を持つイギリス，ブラジル，日本の配信者を選定しており，これは観戦スタイルの差異を検証する上で最適なケーススタディとなる．取得したデータに対し，文脈ごとの意味理解に優れた BERT を用いた感情分析，および短文のクラスタリングに適した BERTopic を用いたトピック分析を適用した．これにより，スラングや口語表現を多く含む SNS テキストからでも高精度な分析が可能となる．そして，短期間でコメント数の多い時間帯でのトピックを時系列で可視化することで言語，地域別の観点から観戦スタイルや文化的差異を明らかにすることを目的とする．

2　関連研究

これまでのスポーツ分野の感情分析は，Twitter(現 X) などのテキストデータを中心に行われてきた．Yu ら [2] は辞書ベースや SVM を用いて 2014 年ワールドカップの分析を行った．この研究では，国ごとのポジティブ・ネガティブの割合を円グラフで提示し，試合の勝敗結果によってサポーターの感情が変化するといった基礎的な相関関係を可視化した．また，試合展開と感情の連動性に着目した研究として，Borg ら [3] はイギリス・プレミアリーグの試合中の Twitter データを分析した．辞書ベースの手法である VADER を用いて，ツイートの感情スコアの変動と，ゴールやファウルといった試合イベントとの相関関係を明らかにした．しかしながら，これらの研究手法は単語の出現頻度や辞書のマッチングに依存するため，文脈を含めた深い意味理解は困難であるという課題がある．例えば，日本語の「やばい」のように，文脈によってポジティブにもネガティ

ブにも解釈される多義的な単語が含まれる場合，その判別精度は著しく低下してしまう．

これらの研究に対して，動画プラットフォーム上のコメントを対象とした研究も行われている．Gil-Lopez ら [4] は，レアル・マドリード対バルセロナの試合に関する YouTube 動画のコメントを分析した．具体的には，英語圏とスペイン語圏のファンを比較し，話題のトピックや感情表現における観戦スタイルの文化的差異を明らかにした．しかし，彼らの研究は動画全体に対するコメントを集合として捉える静的な分析であり，得点シーンなどの試合展開に連動した詳細な時系列分析ではない．

一方, 近年技術面では BERT などの深層学習モデルの登場により, 文脈理解の精度が飛躍的に向上した．実際, Chumwatana [5] は東京オリンピックのツイートを対象に感情分析を行った．この研究では，SVM といった機械学習と BERT といった深層学習モデルの比較実験を実施した．その結果，BERT の方が顕著に分析精度が高かったことを示している．

また，トピックモデリングにおいても，従来の LDA に代わり，BERT の埋め込み表現を利用した BERTopic [6] が注目されている．これらの技術は，オンラインレビュー，ニュース記事の自動分類といった静的なテキストデータの分析において高い成果を上げている．

しかし，これらをスポーツのライブ配信コメントのような，リアルタイムかつ多言語が混在し，文脈が激しく変化するデータに適用した事例はまだ少ない．特に Twitter を用いた研究とは異なり，YouTube Live のコメントは試合映像と完全に同期しており，視聴者が同じ瞬間を共有する「同時性」が高いという特徴がある．さらに，既存研究の多くはハッシュタグ等で抽出された不特定多数のコメントを分析対象としているが，特定の配信チャンネルごとのコメントを分析することで，ホーム・アウェイ・中立といった「明確に異なる立場」からの反応を比較検証した研究は十分に行われていない．

3 分析手法

本研究では，SNS を通じたスポーツ観戦の代表的なプラットフォームの一つである Youtube ライブ配信のコメントデータを分析に用いた．そのデータに対して前処理を行い，BERT による感情分析，および BERTopic によるトピック分析を行った．さらに，国・言語ごとの視聴者の感情覚醒度および評価語の使用傾向を定量的に分析した．

3.1 データ収集

今回，分析に用いた対象データは 2025 年 10 月 14 日に開催された「キリンチャレンジカップ 日本対ブラジル」の Youtube ライブ配信のチャットログである．この試合は，世界的な注目度が高く，試合内容が前半と後半で著しく変化したため，明確な文化・視点の違いが期待できるため良質なデータと考えた．データ取得には，「YouTube Data API v3」を使用し，コメント本文に加え，投稿日時，ユーザ ID を取得した．さらに，コメントは日本，ブラジル，イギリスの計 3 人の配信者から取得しており，言語はそれぞれ，日本語，ポルトガル語，英語である．データ取得期間は試合開始前から試合終了後のまでの約 3 時間に及び，最終的に合計 10,000 件のコメントを分析対象とした．言語別の内訳として，

- 日本語：約 4900 件
- ポルトガル語：約 3200 件
- 英語：約 1900 件

となった．

3.2 データ前処理

本研究では，コメントデータを分析する前に，以下の手順で前処理を行った．

1. ノイズ除去

   まず，Python の正規表現ライブラリであり「re」を使用し，分析に不要な要素である URL，ユーザのメンションを削除した．これにより，テキストの文脈に焦点を当てた分析が可能となった．

2. 言語判定

   配信チャンネルの言語とコメント言語が必ずしも一致しないため，言語判定ツール「langdetect」を使用し，各コメントの言語を正確に分類した．このツールは N-gram ベースの確率的言語判定手法を用いており，短文においても高精度な判定が可能である．判定結果に基づき，日本語，ポルトガル語，英語の 3 つに分類し，それ以外の言語のコメントは分析対象から削除した．

これらの前処理の結果，元データ約 11,500 件から 10,000 件の高品質なコメントデータが得られ，以降の分析に使用した．

3.3 BERT を用いた感情分析

本研究では，各言語のコメントに対して感情分析モデルである XLM-RoBERTa を使用して感情分類を行った．このモデルは，約 2 億件の Twitter 投稿 (現 X) を用いて事前学習されたや言語感情分析モデルであり，100 言語以上に対応している．本モデルは短文かる多言語のソーシャルメディアデータに特化しており，YouTube ライブチャットのような非公式な表現や口語体に対しても高い精度を誇っている．また，このモデルは各コメントを以下のポジティブ・ニュートラル・ネガティブの 3 つの感情カテゴリに分類し，それぞれの確率スコアを出力する．本研究では，最も高い確率を示したカテゴリをそのコメントの感情ラベルとして採用した．

3.4 BERTopic を用いたトピック分析

本研究では，Grootendorst(2022) [6] が提案した，BERTopic ライブラリを用いて，各言語・各フェーズにおけるトピック抽出を行った．BERTopic は，BERT ベースの埋め込み，次元削減，密度ベースクラスタリング，特徴後抽出の 4 段階から構成される細心のトピックモデリング手法である．

1. BERT 埋め込み生成

   まず，Sentence Transformers の多言語モデル「paraphrase-multilingual-MiniLM-L12-v2」を使用し，各

コメントを 384 次元の密ベクトルに変換した．このモデルは 50 以上の言語に対応しており，異なる言語間でも意味的に類似したコメントが近い位置に配置される特性を持つ．

2. UMAP 次元削減

   次に，UMAP により，先ほどの密ベクトルを 5 次元に削減した．UMAP のハイパーパラメータは以下の通り設定した．

   - 削減後の次元数：5
   - 局所構造の保存範囲：15
   - クラスタの密集度:0
   - 距離尺度：コサイン類似度

   これにより，意味的に類似したコメントが近接する 5 次元空間が構築された．

3. HDBSCAN 密度ベースクラスタリング

   削減された 5 次元空間に対して，HBDSCAN によりクラスタリングを行った．HDBSCAN は密度の高い領域を自動的にクラスタとして検出し，トピック数を事前に指定する必要がないという利点をもつ．設定パラメータは以下のとおりである．

   - クラスタとして認識する最小サイズ：20
   - コアポイント判定の閾値：5
   - 距離尺度：ユークリッド距離

   この手法により，各言語・各フェーズにおいて自動的にトピック数が決定された．

4. c-TF-IDF 特徴後抽出

   最後に，この指標を用いることで，各トピックの特徴的なキーワードを抽出した．c-TF-IDF は，通常の TF-IDF をトピックレベルに拡張したものであり，各トピック内で頻出かつ他トピックでは稀な単語をハイスコアとして抽出する．今回は上位 5-10 ワードを各トピックの代表語として採用した．

3.5 時系列分析とバースト地点検出

視聴者の反応の変化を時系列で詳細に分析するため，各フェーズにおいてコメント数が急増する「バースト地点」を検出し，その地点でトピック内容を可視化した．まず，5 分間隔でコメント数を集計し，各フェーズにおいて最もコメント数の多い上位 3 地点をバースト地点として抽出した．ただし，短時間で複数のバースト地点が集中してしまう現象を避けるため，各バースト地点は最低 10 分以上の間隔になるように制約を設けた．次に，抽出された各バースト地点において，検出地点を中心としたコメント群を対象に BERTopic 分析を実施した．また，今回はバースト地点で抽出されたコメントのうち，試合寧ように関係ないものを除外するため，多言語キーワードベースのフィルタリングを行った．最後に，各バースト地点において，視聴者の主要な反応を示す代表コメントを表示した．これらの過程により，全 27 バースト地点におけるトピック内容を可視化し，時系列での視聴者反応の変化を定量的に分析することが可能となった．時系列グラフ上にバースト地点を赤色の点で示し，代表コメントとトピックワードを黄色の吹き出しで表示した．

各フェーズにおいて，5 分間隔でコメント数をカウントし，最もコメント数の多い上位 3 地点をバースト地点として抽出した．各バースト地点において，検出地点± 5 分間のコメントを対象に BERTopic トピック分析を実施した．この手法により，全 27 バースト地点において確実にトピック内容を可視化し，時系列での視聴者反応の変化を定量的に分析することが可能となった．

3.6 覚醒度スコアの定義

視聴者のコメントに含まれる感情的な高まりを定量化するため，本研究では NRC VAD Lexicon [7] および VADER [8] の設計思想を参考にし，コメント中の語彙的覚醒度と強調表現を統合したスコアを定義した．ただし，本指標は心理学的な arousal を直接推定するものではなく，コメントにおける感情的強調度の代理指標として解釈すべきである．

算出プロセスは以下の通りである．

1. **ベーススコア（語彙的覚醒度）**：Mohammad (2018) による NRC VAD Lexicon を用い，各コメントに含まれる単語の覚醒度（0.0〜1.0）を取得した．短文テキストでは感情表現が局所的に出現しやすい点を考慮し，コメント内で辞書にマッチした単語の覚醒度の最大値を $\mathrm{VAD}_{\max}$ として採用した．
2. **強調表現による補正**：Hutto & Gilbert (2014) の VADER における強調表現の扱いに基づき，感嘆符の連続使用および All-Caps を強調の手掛かりとして加算した．具体的には，感嘆符数 $n_!$（最大 4）に対して $0.292 \times n_!$ を加算し，All-Caps が一定割合を超える場合に 0.733 を加算した．All-Caps は日本語に適用不可能であるため，言語間比較においては測定バイアスとなり得る．

以上より，本研究の強調度スコアは次式で定義する．

$$\text{Arousal Score} = \min\left(1.0,\ \mathrm{VAD}_{\max} + 0.292 \times n_! + 0.733 \times \mathbb{1}_{\mathrm{CAPS}}\right) \tag{1}$$

ここで，$\mathbb{1}_{\mathrm{CAPS}}$ は All-Caps に該当する単語が一定割合を超える場合に 1，そうでない場合に 0 を取る指示関数である．また，最終値を $[0, 1]$ にクリッピングするため，高強調コメントでは天井効果が生じ得る．

なお，絵文字については既存研究 [9] を参考に補助的特徴として扱ったが，絵文字から arousal を直接推定することは困難であるため，本稿の「覚醒度スコア」定義式（式 1）には含めず，結果解釈においては別要因として扱う．

3.7 評価語比率

覚醒度に加え，視聴者が試合内容に対してどの程度「分析的・評価的」な言語表現を用いているかを把握するため，「評価語比率（Evaluation Ratio）」を定義した．算出にあたっては，LIWC [10] のカテゴリ設計の考え方を参考にしつつ，本研究の対象（スポーツ観戦チャット）に合わせて語彙リストを構築した．ただし，標準化された多言語 LIWC 辞書を直接利用したものではないため，本指標は評価語彙の出現頻度の代理指標として解釈すべきである．

評価語比率は，以下の 2 カテゴリに属する単語の総数を，コメント内の全実質語数（Total Content Words）で除算することで算出した（式 2）．

$$\text{Evaluation Ratio} = \frac{\text{Cognitive Words} + \text{Judgment Words}}{\text{Total Content Words}} \quad (2)$$

1. **認知的プロセス語（Cognitive）**：LIWC の Cognitive Processes のカテゴリ設計を参考に，「洞察（思う，考える）」，「因果（だから，なぜなら）」，「推量（たぶん，かもしれない）」に該当する語彙を各言語で整理した．
2. **判断語（Judgment）**：ドメイン依存語彙を辞書として構築するアプローチ [11] を参考に，サッカー観戦における質の判定（上手い，下手，雑，good，bad 等）に相当する語彙を各言語で収集・整理した．戦術用語については評価語と混同しやすいため，本研究では判断語とは分離し，必要に応じて別途の分析対象とした．

なお，否定表現（例：良くない，not good）については簡易ルールにより検出を試みたが，短文・口語・反語・二重否定では誤判定が生じ得る．このため，本稿の評価語比率は，厳密な「評価行動」ではなく，評価語彙の使用傾向を表す指標として記述的に扱う．

3.8 測定の限界

本研究で使用した覚醒度・評価語比率の算出には，多言語比較における測定等価性の観点から以下の制約がある．

1. 第一に，辞書カバレッジの言語間差異が大きい．NRC VAD Lexicon ベースの語彙でのトークンレベルヒット率は，英語 1.11%，ポルトガル語 0.51%，日本語 0.09%であり，日本語は英語の約 1/12 に留まる．これは使用した辞書が英語中心（約 20,000 語）であるのに対し，日本語・ポルトガル語は研究者が補助的に追加した数十語のみであることに起因する．
2. 第二に，VADER の全大文字ルールは日本語に適用不可能であり（ひらがな・カタカナに大小文字の概念がない），言語バイアスが生じる．
3. 第三に，評価語辞書は標準化された多言語 LIWC 辞書に基づいておらず，研究者が構築したものである（英語 13 語，日本語 8 語，ポルトガル語 7 語の認知語を使用）．

これらの制約により，本研究の結果は視聴スタイルの差異だけでなく，測定バイアスを含む可能性があり，解釈には慎重を要する．また，本稿で導入した両指標は厳密な心理尺度ではなく，ライブチャットにおける言語行動の特徴量として位置づけ，記述的比較を主とする．

4 実験結果

本章では，実験結果に関して 4.1 で BERT による感情分類を用いた時系列感情推移と，前半，後半，試合後の 3 つのフェーズの各バースト地点の BERTopic を用いた時系列トピック抽出について，4.2 で 3 か国にごとのコメントの覚醒度および評価語の比率について，それぞれの結果とそれに対する考察をしていく．

4.1 時系列トピック抽出

まず，図 1 に日本，ブラジル，イギリス 3 ヵ国の前半・後半・試合後の 3 つのフェーズでの感情推移をポジティブ・ニュートラル・ネガティブの 3 つに分けて可視化したものを表す．図の上部はポジティブ領域，中央部はニュートラル，下部はネガティブ領域を表している．

図 1 より，日本について，前半は 2 点のリードを許しているためネガティブになっているが，下馬評通りのリードされている展開なため大きくネガティブになっているわけではない．しかし，後半は，日本が逆転したため，感情もネガティブに推移していることが分かる．次に，ブラジルは前半に 2 点リードし日本を圧倒していたため，ポジティブ領域に感情が推移していることが分かる．下馬評通りの試合展開のため，過度にポジティブになっていないと考えられる．しかし，後半は日本に逆転され，予想外であったため感情が大きくネガティブに推移していることが分かる．最後に，イギリスについて，前半・後半・試合後どのフェーズでもブラジルと日本の間に推移しており，どの時間帯であっても中立的な立場でコメントをしていることが分かる．

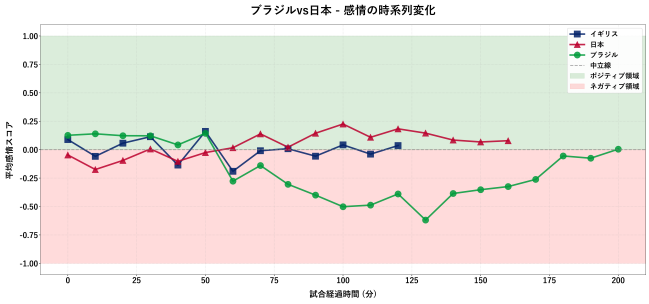


図 1　各国の全フェーズの感情推移

次に，図 2，図 3，図 4 に日本の前半・後半・試合後の 3 つのフェーズ内における各バースト地点を BERTopic でトピック抽出した結果を示す．その次に，図 5，図 6，図 7 にブラジルの前半・後半・試合後の 3 つのフェーズ内における 3 つのバースト地点を BERTopic でトピック抽出した結果を示す．最後に，図 8，図 9，図 10 に日本の前半・後半・試合後の 3 つのフェーズ内における各バースト地点を BERTopic でトピック抽出した結果を示す．

図 2 より，日本は負けているため，抽出されたトピックは試合状況を批判する，試合時間 30 分頃の「アンチェロッティさん流石の崩し」といった対戦相手であるブラジルの監督を賞賛するトピックが抽出されていることが分かる．図 3 より，日本は逆転しているのにもかかわらず，自国のチームを賞賛するポジティブなトピックが抽出されていない．また，試合時間 75 分頃には「サブ組もやばいなブラジル」といったブラジルを揶揄するようなトピックがで盛り上がっていることが分かる．さらに，図 4 より，ブラジルの選手を心配するトピックや，自国の

監督について評価しているトピックが抽出されており，今後の展望で盛り上がっていることが分かる．

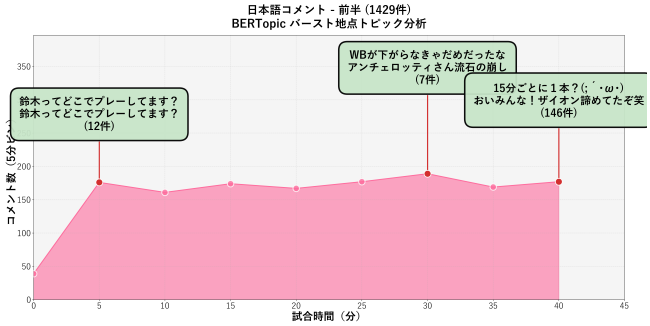


図 2　日本の前半のバースト地点でのトピック推移

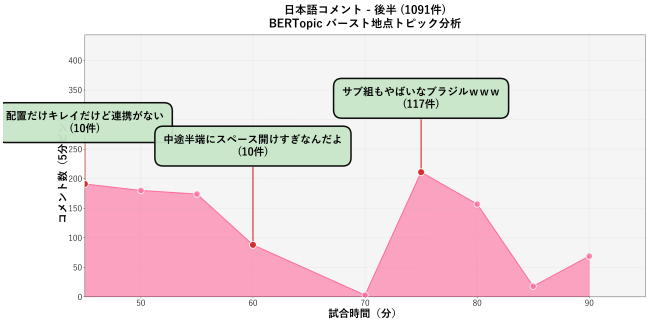


図 3　日本の後半のバースト地点でのトピック推移

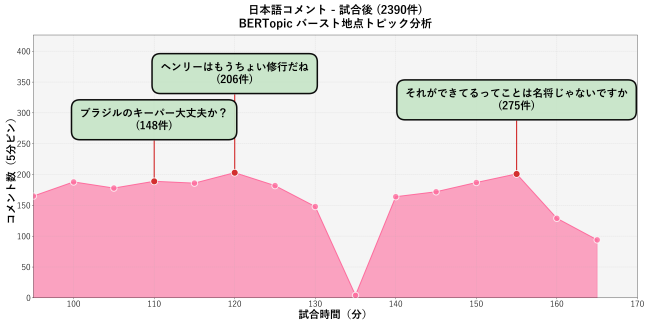


図 4　日本の試合後のバースト地点でのトピック推移

ブラジルについて，図 5 より，ブラジルのサポーターは試合時間 5 分頃の「3-1 でブラジルが勝つ」や 40 分ごろの「ヴィニシウス行け」といったポジティブなトピックで盛り上がっている．次に，図 6 より，試合時間 50 分頃に「ゴールを決めろ」といった鼓舞するトピックが抽出されたのに対し，直後に日本にゴールを決められた 55 分頃には「ネイマールのいない代表チームなんて面白くない」といったネガティブなトピックが抽出されている．さらに，75 分頃には「やっぱりナルトだったか」トピックが抽出されており，日本代表チームのことを人気漫画「ナルト」に比喩，90 分頃には「信じられない」といったトピックが抽出されており，ブラジルの視聴者は負けを認めつつも対戦相手を賞賛していることが分かる．その次に，7 より，110 分頃に「俺たちがドイツじゃなくて運が良かった」というトピックが分類されており，日本がカタールワールドカップでドイツを破ったことをなぞらえたジョークで盛り上がっていることが分かる．また，125 分頃には漫画「ナルト」の登場人物である「ジライヤ」というキャラクターを日本代表に例えて賞賛しているトピックで盛り上がっており，ブラジルサポーターが日本の強さを認めていることが分かる．

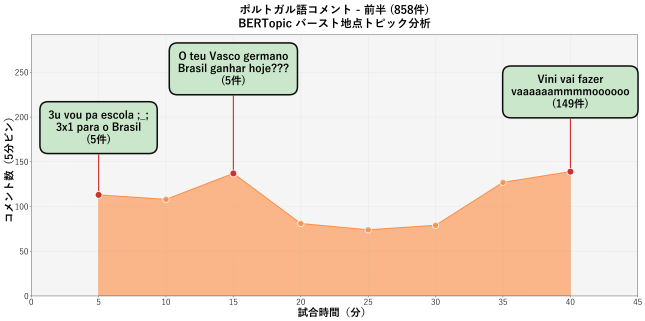


図 5　ブラジルの前半のバースト地点でのトピック推移

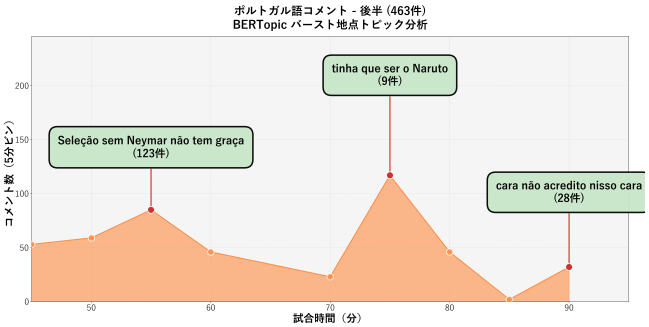


図 6　ブラジルの後半のバースト地点でのトピック推移

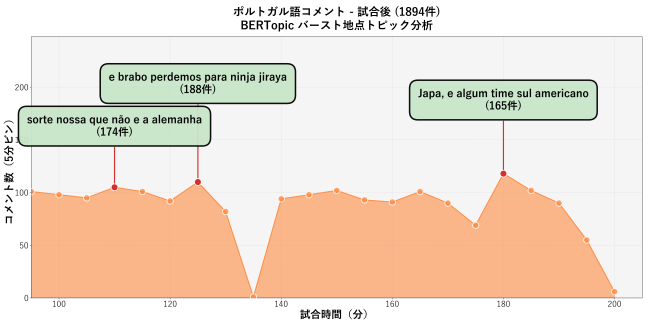


図 7　ブラジルの試合後のバースト地点でのトピック推移

イギリスについて，図 8 より，試合時間 30 分頃に「最初に流れを作ったのはマルティネッリ，堂安律」と第三者視点から試合を分析していることが分かる．また，図 9 より，試合時間 90 分頃に「ブラジルの守備がひどい」というトピックが抽出されており，逆転をされたブラジルに対しての評価をしていることが分かる．最後に，図 10 より，95 分頃にブラジルのチームに対する批判であったり，120 分頃には「ジョエリントン」という途中出場した選手について批判しているトピックが抽出された．

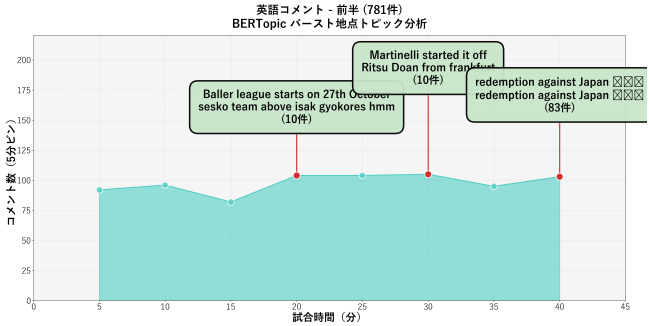


図 8　イギリスの前半のバースト地点でのトピック推移

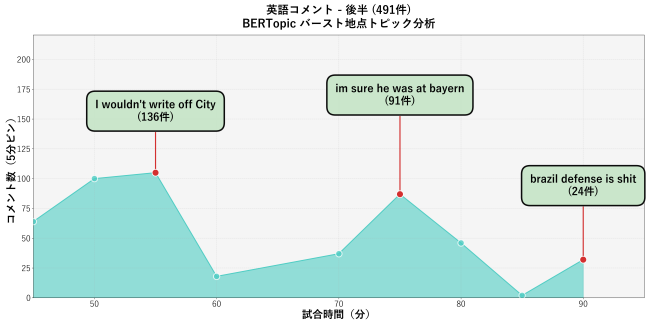


図 9　イギリスの後半のバースト地点でのトピック推移

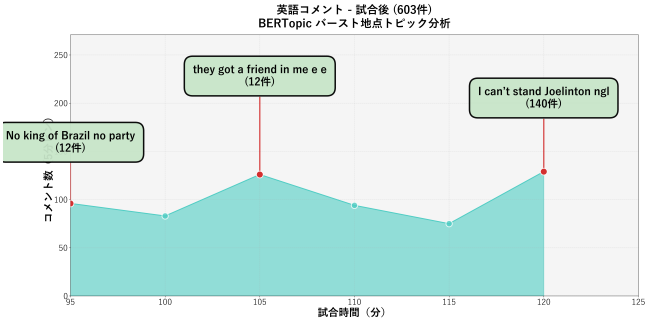


図 10　イギリスの試合後のバースト地点でのトピック推移

4.2　覚醒度と評価語比率分析

各国の観戦スタイルの質的な違いを明らかにするため，算出した「平均覚醒度」と「評価語比率」の 2 軸で各国の位置づけを可視化した（図 11）．

図 11 を参照すると，各国はそれぞれ異なる象限に位置しており，明確な特性の違いが確認された．

1. **ブラジル（黄色）：熱狂的感情表出型**
   グラフ右側に位置し，覚醒度が他国と比較して顕著に高い．評価語比率も比較的高水準である．これは，プレーに対して「VAMOS」などの絶叫や強調表現を多用し，感情の爆発とプレーへの反応が直結している「熱狂的」なスタイルを示している．
2. **イギリス（青色）：バランス型観戦**
   グラフ中央付近に位置する．覚醒度・評価語比率ともに中程度であり，第三者視点として適度な興奮と冷静な分析を併せ持つ，バランスの取れた観戦態度がうかがえる．
3. **日本（赤色）：静観的観察型**
   グラフ左下に位置し，覚醒度・評価語比率ともに最も低い値を示した．これは，従来の定性的な印象である「日本人は大人しい」という傾向を定量的に裏付けるものである．チャット上での感情表出（覚醒）を極力抑え，かつ積極的な言語化（評価）も行わずに試合を見守る「静観的」なスタイルが，日本の YouTube 観戦文化の特徴として示唆された．

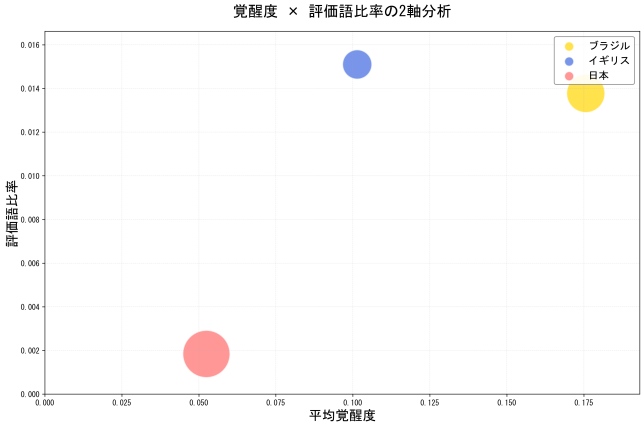


図 11　各国の平均覚醒度と評価語比率の散布図

5　まとめと今後の課題

本研究では，近年普及している SNS を通じたリアルタイムスポーツ観戦に着目し，YouTube ライブ配信のコメントデータを用いて，国・言語別（日本語・ポルトガル語・英語）の応援スタイルおよび感情変化の差異を分析した．分析手法として，まず文脈理解に優れた BERT を用いた感情分析と，BERTopic を用いた時系列トピックモデリングを適用し，試合展開（前半・後半・試合後）に応じた視聴者の反応を可視化した．さらに，スポーツ観戦における視聴者の心理状態をより多角的に捉えるため，従来の感情極性（ポジティブ・ネガティブ）に加え，「覚醒度」および「評価語比率」という二つの新たな定量的指標を導入し，分析を行った．

2025 年に行われたサッカー「日本対ブラジル」戦を対象とした実験の結果，以下の知見が得られた．

1. **日本：** 試合に勝利（逆転）している状況下でも，試合内容への批判や反省点などのネガティブなトピックが抽出される傾向が見られた．これは，目の前の勝敗よりも内容や将来の展望を重視する，日本特有の批判的かつ悲観的な応援スタイルを示唆している．
2. **ブラジル：** 感情の振幅が大きく，優勢時は強くポジティブに，劣勢時は強くネガティブに反応する傾向があった．また，「ナルト」や「忍者」といった日本のポップカルチャーを比喩に用いて相手チーム（日本）を称賛するなど，ユーモアとリスペクトを含んだ独自の表現文化が明らかになった．
3. **イギリス（第三者視点）：** 当事国に比べて感情の起伏が小さく，試合についてのトピックはありつつも，試合展開と

は直接関係のない話題も見られた．結果として，エンターテインメントとして冷静に，あるいは俯瞰的に観戦するスタイルが確認された．

以上のトピック分析の結果より，本研究の手法は，単なる勝敗への反応だけでなく，各言語圏に根付く文化的背景や観戦スタイルの「質的な違い」を時系列で可視化できることを示した．

次に，視聴者の「覚醒度」および「評価語比率」を，NRC VAD Lexicon や VADER，LIWC といった心理言語学の標準的手法を用いて算出し，国・言語別の観戦スタイルに関する定量的差異を明らかにした．

まず，国ごとの「熱狂」と「批評」のバランスにおいて顕著な違いが確認された．ブラジルの視聴者は，他国と比較して極めて高い平均覚醒度と，比較的高い評価語比率を示した．これは，特定のプレーに対して記号や強調表現（大文字，感嘆符の連打）を多用し，感情的な高まりとプレーへのジャッジをストレートにチャット上に投影する熱狂的かつ参加型の観戦スタイルを裏付けている．対照的に，日本の視聴者は覚醒度・評価語比率ともに最も低い数値を示した．これは，感情的な興奮（絶叫）を抑制し，かつチャット上での積極的な言語化も行わずに淡々と試合を見守る静観的なスタイルが定着していることを定量的に示唆している．また，第三者視点である英語圏の視聴者は，覚醒度・評価語比率ともに中程度に位置し，適度な興奮と冷静な分析を併せ持つバランスの取れた傾向が示された．

加えて，感情的な「盛り上がり」と「評価行動」の独立性が示された．相関分析の結果，覚醒度と評価語比率の間には有意な相関が見られず，これらが独立した次元であることが確認された．この発見は，従来の単一的な感情分析だけでは，単に騒いでいるのか，冷静に分析しているのか，あるいは静観しているのかといった，スポーツ観戦体験の質的な多様性を完全には捉えきれないことを実証するものである．

以上の結果より，本研究で構築した，恣意性を排除し構文解析を取り入れた分析フレームワークは，言語や文化的背景によって異なる観戦行動の質的差異を比較可能な形で可視化する上で有効であると結論付けられる．

一方，本研究では YouTube ライブ配信のコメントという新しいデータソースを用いた多言語間の観戦スタイル比較において一定の成果を得たが，いくつかの課題も残された．

第一に，データセットの拡張である．本研究では「日本対ブラジル」という特定の 1 試合のみを対象とした．そのため，今回明らかになった「日本人の悲観的な応援傾向」や「ブラジル人の比喩的な表現」が，この試合特有のものなのか，あるいは国民性として一般化できるものなのかを検証するにはデータ数が不十分である．今後は，さらにワールドカップといった，注目度の高い国際試合や競技（野球やバスケットボールなど）へと対象を広げ，分析結果の普遍性を高める必要がある．

第二に，第三者視点（英語圏）のデータ品質の向上である．実験結果（図 8〜10）において，英語コメントからは試合展開と直接関係の薄いトピックが多く抽出された．これは，選択した配信チャンネルの視聴者層が試合そのものよりもチャット内での雑談を重視していた可能性や，言語判定の精度によるノイズが含まれていた可能性が考えられる．より純粋な「第三者視点の観戦スタイル」を抽出するためには，チャンネル選定基準の厳格化や，試合関連用語に基づくフィルタリング処理の導入が求められる．

第三に，マルチモーダル分析への展開である．現在はテキストデータのみを対象としているが，YouTube には映像と音声が含まれている．得点シーンや反則シーンなどの映像特徴量とコメントのバーストを統合的に分析することで，どのようなプレーが各国の感情を大きく動かしたのかを，より高い解像度で特定できると考えられる．

第四に，多言語比較における測定等価性の検証である．本研究では，YouTube Live チャットという新しいデータソースから視聴者の感情・認知パターンを定量化する試みとして，NRC VAD Lexicon や VADER，LIWC といった心理言語学の標準的手法を適用した．しかし，これらの辞書は本来英語を対象として開発されたものであり，多言語環境における適用には辞書カバレッジの不均衡という課題が存在する．本研究の検証によれば，日本語コメントにおける VAD ヒット率は英語の約 1/12 に留まっており，言語間での測定等価性が十分に確保されていない可能性がある．この問題は，各言語圏の観戦スタイルの「真の差異」と「測定手法の言語バイアス」を区別することを困難にする．今後の研究では，多言語 VAD 辞書の拡充，翻訳ベースアプローチの併用，または多言語 BERT モデルによるエンドツーエンド感情推定の導入により，構成概念妥当性を高めた上での言語間比較を行う必要がある．このような方法論的改善を通じて，より信頼性の高い多言語感情分析基盤を構築することが求められる．

以上の課題に取り組むことで，スポーツ観戦における文化横断的な感情表現の理解がさらに深まり，SNS 時代のスポーツマーケティングや国際コミュニケーション研究への応用が期待される．

文　献

[1] 笹川スポーツ財団. スポーツライフ・データ（スポーツ観戦率の推移）. https://www.ssf.or.jp/thinktank/sports_life/data/sportsspectating.html, 2024. (参照 2025-12-21).

[2] Yan Yu and Xiaohua Wang. World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans. *Computers in Human Behavior*, Vol. 48, pp. 392–400, 2015.

[3] Adam Borg and Martin Boldt. Using vader sentiment analysis to match word-of-mouth with game events in the english premier league. *Systems*, Vol. 8, No. 4, p. 48, 2020.

[4] Teresa Gil-Lopez, Tang Li, Viorela KD, et al. One ball, two worlds: A comparative analysis of the ‘el clásico’ fans’ commenting behavior on youtube. *International Journal of Sport Communication*, Vol. 11, No. 1, pp. 67–87, 2018.

[5] Tanes Chumwatana. Twitter sentiment analysis for the 2020 tokyo olympic games using deep learning. *International Journal of Cloud Applications and Computing (IJCAC)*, Vol. 12, No. 1, pp. 1–17, 2022.

[6] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[7] Saif M. Mohammad. Obtaining reliable human ratings of va-

lence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1, pp. 174–184, 2018.

[8] C. J. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, pp. 216–225, 2014.

[9] Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PLoS ONE*, Vol. 10, No. 12, p. e0144296, 2015.

[10] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of LIWC2015. Technical report, University of Texas at Austin, 2015.

[11] Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, Vol. 66, No. 1, pp. 35–65, 2011.

自動話題分割とコメント分析による動画セグメントの人気度推定

小柳津海来† 石橋 直樹†

† 武蔵野大学データサイエンス学部データサイエンス学科 〒135–8181 東京都江東区有明 3 丁目 3 番 3 号
E-mail: †s2422008@stu.musashino-u.ac.jp, ††n-ishi@musashino-u.ac.jp

あらまし 本研究では，動画を話題単位で自動分割し，各話題の人気度をコメント分析により測定する手法を提案する．Sentence Transformer と PELT 法により字幕から話題境界を検出し，タイムスタンプ付きコメントから 4 つの指標（コメント密度，感情スコア，ポジティブ率，感情一致度）を算出して人気度スコアを計算する．YouTube Live 配信動画 (約 54 分) を用いた実験では，6 個のセグメントに分割され，人気度スコア 30.56–65.83 の範囲で算出された．本手法により，視聴者に支持される話題の自動特定が可能となり，コンテンツ制作支援や視聴者の効率的な動画消費への応用が期待される．

キーワード 話題セグメンテーション，動画コンテンツ分析，視聴者行動分析，感情分析，YouTube

1 はじめに

近年，YouTube をはじめとする動画共有プラットフォームの利用が拡大している．特にゲーム実況，ライブ配信，教育動画などのジャンルでは，1 時間を超える長時間動画が一般的となっている．しかし，動画の長時間化に伴い，視聴者が興味のある部分を見つけることは困難になっている．既存のチャプター機能は投稿者の手動設定に依存し，すべての動画で提供されているわけではない．また，視聴者が実際に興味を持つ部分と，投稿者が設定したチャプターが必ずしも一致しないという問題もある．

一方，技術面では音声認識，自然言語処理，感情分析の各分野で大きな進展があった．YouTube では自動字幕生成機能により多言語の字幕が利用可能であり，またライブ配信ではチャットリプレイ機能によりタイムスタンプ付きコメントデータが取得できる．これらのデータを活用することで，視聴者の反応を定量的に分析することが可能となっている．

既存研究では，動画の自動セグメンテーションとコメント分析が個別に扱われてきた．動画セグメンテーション研究は主に視覚情報に基づくショット検出が中心であり，話題の意味的変化を捉えることは困難である．コメント分析研究も動画全体を対象としたものが多く，話題単位での詳細な分析は行われていない．

本研究では，これらの技術を統合し，動画を話題単位で自動分割した上で，各話題の人気度をコメント分析により定量的に測定する手法を提案する．これにより，視聴者に支持される話題の特徴を明らかにし，コンテンツ制作者への有益なフィードバック提供や，視聴者の効率的なコンテンツ消費支援を実現する．

2 関連研究

動画コンテンツ分析とコメント分析に関する研究は，これまで個別に扱われてきた．

堺ら [3] は YouTube における視聴者コメントの特徴抽出を行い，炎上動画の自動検出を試みている．彼らは感情分析 API を用いてコメントをベクトル化し，機械学習により炎上動画を分類する手法を提案した．本研究との違いは，堺らが動画全体のコメント傾向から炎上を検出するのに対し，本研究では動画を話題単位に分割し，各セグメントの人気度を測定する点である．

動画セグメンテーションに関しては，視覚情報に基づくショット検出や音声認識を用いた手法が主流である．しかし，これらの手法は映像の物理的な切り替わりは検出できるものの，話題の意味的変化を捉えることは困難である．本研究では，Sentence Transformer と PELT 法を組み合わせることで，字幕テキストの意味的な変化点を検出し，話題境界を特定する．

コメント分析に関しては，感情分析や視聴者行動分析の研究が行われているが，動画全体を対象としたものが多く，話題単位での詳細な分析は行われていない．本研究では，セグメント単位でコメントを分析することで，どの話題が視聴者に支持されているかを定量的に測定できる点が新規性である．

3 提案手法

3.1 システム概要

提案システムは以下の 5 つの処理フェーズから構成される: (1) YouTube 字幕データの取得と前処理，(2) テキストベースの話題セグメンテーション，(3) チャットリプレイからのコメントデータ取得，(4) コメントデータの感情分析，(5) 各セグメントの人気度スコア算出．図 1 に提案システムの処理フローを示す．

3.2 字幕データの取得と前処理

YouTube の自動生成字幕 API を用いて，動画の字幕データを取得する．取得した字幕は短い時間間隔で分割されているため，前処理として時間的に近接する字幕を統合する．統合ウィンドウサイズを w 秒として，時刻 t から $t+w$ までの字幕を 1 つのテキストチャンクにまとめる．これにより，セグメンテー

ションに適した粒度のテキストデータを得る．

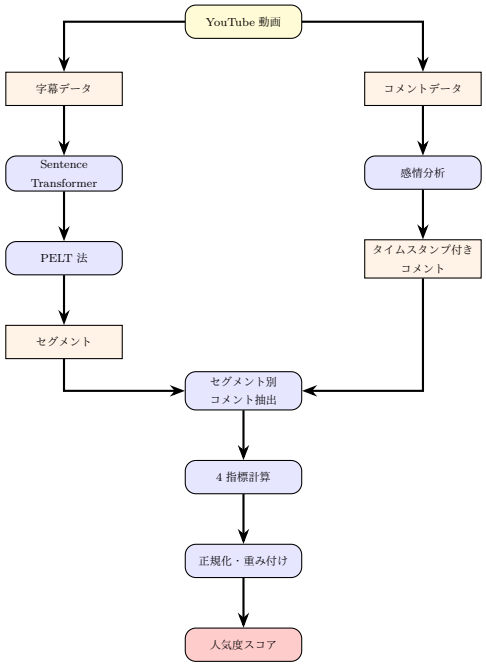


図 1　提案システムの処理フロー

3.3　話題セグメンテーション

統合されたテキストチャンクに対し，Sentence Transformer [1] を用いて埋め込みベクトルに変換する．本研究では “paraphrase-multilingual-MiniLM-L12-v2” モデルを使用し，384 次元のベクトル表現を得る．

得られた埋め込みベクトル系列 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ に対し，PELT (Pruned Exact Linear Time) 法 [2] を用いて変化点を検出する．隣接するベクトル間のコサイン距離を計算し，距離系列に対して PELT 法を適用する．変化点は話題境界とみなされ，これにより動画は意味的に一貫した複数のセグメントに分割される．

セグメンテーションでは，閾値 θ と最小セグメント長 $l_{\min}$ の 2 つのパラメータを調整することで，分割の粗さを制御できる．

3.4　コメントデータの取得と処理

YouTube Live 配信のチャットリプレイデータを yt-dlp ツールを用いて取得する．各コメントには投稿者名，テキスト，タイムスタンプが含まれる．タイムスタンプは動画開始からの相対時刻 (秒) として正規化される．

取得したコメントデータに対し，キーワードベースの簡易感情分析を適用する．ポジティブキーワード集合 P とネガティブキーワード集合 N を用意し，各コメント c に含まれるキーワード数 $|c \cap P|$, $|c \cap N|$ をカウントする．感情スコア $s(c)$ を以下の式で計算する:

$$s(c) = \frac{|c \cap P| - |c \cap N|}{|c \cap P| + |c \cap N| + 1} \tag{1}$$

この値は -1(完全ネガティブ) から $+1$(完全ポジティブ) の範囲をとる．

3.5　人気度スコアの算出

各セグメント S_i に対し，タイムスタンプが S_i の時間範囲内に含まれるコメント集合 C_i を抽出する．以下の 4 つの指標を計算する:

コメント密度 D_i:　セグメントの長さ (秒) あたりのコメント数

$$D_i = \frac{|C_i|}{t_{\text{end}}^{(i)} - t_{\text{start}}^{(i)}} \tag{2}$$

平均感情スコア $\bar{S}_i$:　セグメント内コメントの平均感情スコア

$$\bar{S}_i = \frac{1}{|C_i|} \sum_{c \in C_i} s(c) \tag{3}$$

ポジティブ率 P_i:　閾値 τ(本研究では 0.3) 以上の感情スコアを持つコメントの割合

$$P_i = \frac{|\{c \in C_i : s(c) \geqq \tau\}|}{|C_i|} \tag{4}$$

感情一致度 U_i:　感情スコアの標準偏差の逆数 (一貫性の指標)

$$U_i = 1 - \sigma(s(C_i)) \tag{5}$$

これらの指標を 0–100 のスケールに正規化し，重み w_D, w_S, w_P, w_U(合計 1) を用いて統合する．セグメント i の人気度スコア H_i は以下のように計算される:

$$H_i = 100 \times (w_D D_i' + w_S \bar{S}_i' + w_P P_i' + w_U U_i') \tag{6}$$

ここで，プライム記号は正規化された値を示す．本研究では $w_D = 0.35$, $w_S = 0.30$, $w_P = 0.20$, $w_U = 0.15$ を使用した．これらの重み設定の意図は以下の通りである: コメント密度 ($w_D = 0.35$) は視聴者の関心の高さを直接的に示す最も客観的な指標であるため最大の重みを設定した．平均感情スコア ($w_S = 0.30$) は視聴者の感情的反応の質を示す重要な指標として 2 番目に大きな重みとした．ポジティブ率 ($w_P = 0.20$) は肯定的な反応の割合を示し，感情スコアを補完する指標として中程度の重みを設定した．感情一致度 ($w_U = 0.15$) は視聴者の反応の一貫性を示す副次的な指標として最小の重みとした．

4　実装と評価

4.1　実験設定

提案手法の有効性を検証するため，YouTube Live の配信アーカイブ動画 (動画 ID: yw7D79B-_x0, 長さ約 54 分) を対象とした実験を行った．この動画を選定した理由は以下の 3 点である: (1) 複数の明確な話題（神社配信，おみくじ企画，絵馬紹介，重大告知，新曲披露等）を含む長時間配信であり，話題セグメンテーションの評価に適している，(2) チャットリプレイデータが豊富（42,530 件）であり，コメント分析に十分なデータ量がある，(3) 公開動画であるため研究の再現性が確保できる．

セグメンテーションパラメータは，統合ウィンドウ $w = 180$ 秒，閾値 $\theta = 0.85$，最小セグメント長 $l_{\min} = 3$ チャンクとした．これらのパラメータは複数の値を試行した結果，経験的に

最適と判断したものである．統合ウィンドウ 180 秒は，約 3 分間の発話内容を 1 つのチャンクとすることで，話題の切り替わりを適切に捉えられる粒度と考えられる．閾値 0.85 は，過度に細かい分割を避けつつ，意味的な変化を検出できる値である．最小セグメント長 3 チャンクは，極端に短いセグメントの生成を防ぐための制約である．

実装は Python 3.10 で行い，Sentence Transformers ライブラリ，ruptures (PELT 実装)，yt-dlp，pandas，matplotlib を使用した．

4.2 セグメンテーション結果

表 1 に得られた 6 個のセグメントの詳細を示す．各セグメントの長さは約 530–551 秒 (平均 544 秒) でほぼ均等であり，極端に短いセグメントは生成されなかった．これは最小セグメント長の制約が有効に機能したことを示している．

表 1　セグメント詳細とメトリクス

ID	開始時刻 (秒)	終了時刻 (秒)	長さ (秒)	コメント数	密度 (件/秒)
0	4.6	552.8	548.2	6,289	11.47
1	548.2	1089.6	541.4	6,380	11.78
2	1095.4	1643.3	547.9	6,980	12.74
3	1640.8	2192.0	551.2	7,227	13.11
4	2188.6	2718.6	530.0	6,515	12.29
5	2737.4	3283.2	545.8	6,833	12.52

コメント密度は 11.47–13.11 件/秒の範囲であり，セグメント 3(重大告知部分) で最大値を示した．これは視聴者の関心が高い話題であることを示唆している．

4.3 人気度スコア分析

表 2 に各セグメントの人気度スコアと内訳を示す．人気度スコアは 30.56(セグメント 0) から 65.83(セグメント 3) の範囲となった．

表 2　セグメント別人気度スコアとメトリクス

ID	密度	感情	ポジ率	一致度	スコア
0	11.44	0.037	0.074	0.981	30.56
1	11.74	0.047	0.096	0.976	36.97
2	12.72	0.139	0.278	0.949	**64.44**
3	13.08	0.049	0.100	0.976	**65.83**
4	12.26	0.037	0.079	0.978	47.24
5	12.48	0.067	0.144	0.963	51.50

セグメント 2(絵馬紹介) とセグメント 3(重大告知) が高い人気度を示した．セグメント 2 は最も高い平均感情スコア (0.139) とポジティブ率 (0.277) を記録しており，視聴者が肯定的な感情を表現していたことがわかる．セグメント 3 は最高のコメント密度 (13.11 件/秒) を示し，視聴者の活発な反応があったことを示している．

図 2 に人気度スコアの時系列変化を示す．グラフから，配信の中盤から後半 (セグメント 2–5) にかけて人気度が上昇する傾向が見られる．これは，配信が進むにつれて視聴者の期待が高まり，クライマックス (重大告知，新曲披露) で盛り上がりを見せる典型的なライブ配信の構造を反映していると考えられる．

図 2　セグメント別人気度スコアの時系列変化

図 3 に各セグメントの人気度スコア内訳を示す．すべてのセグメントでコメント密度の寄与が最も大きく，次いで感情スコア，ポジティブ率の順となっている．

図 3　人気度スコアのメトリクス別内訳

4.4 考　　察

実験結果から，以下の知見が得られた:

1. **セグメンテーションの妥当性:** 統合ウィンドウ 180 秒という長めの設定により，適度な粒度でのセグメント分割が実現された．各セグメントが約 9 分という長さは，配信の主要な話題の切り替わりを捉えるのに適切であったと考えられる．
2. **人気度指標の有効性:** 人気度スコアが高かったセグメント 2，3 は，それぞれ視聴者参加型企画 (絵馬紹介) と重要なお知らせ (ライブ・展示会告知) という，視聴者の関心が高い内容であった．これは提案手法が視聴者の反応を適切に捉えていることを示唆している．
3. **感情分析の限界:** 今回はキーワードベースの簡易的な感情分析を用いたが，感情一致度が全セグメントで 0.95 以上と高く，コメントの感情的多様性を十分に捉えられていない可能性がある．より高度な感情分析モデル (BERT 系モデル等) の導入が今後の課題である．

5 おわりに

本研究では，YouTube 動画を話題単位で自動分割し，各セグメントの人気度をコメント分析により定量的に測定する手法を提案した．実際のライブ配信動画を用いた実験により，提案手法が視聴者の反応を適切に捉え，人気の高い話題を特定できることを示した．

今後の課題として，(1) より高度な感情分析モデルの導入，(2) 多様なジャンルの動画への適用と評価，(3) リアルタイム処理への対応，などが挙げられる．また，本手法を活用したコンテンツ制作支援ツールや，視聴者向けのハイライト自動生成システムの開発も今後の展望である．

謝　　辞

本研究で使用した動画データの提供にご協力いただいた配信者の方に感謝申し上げます．

文　　献

[1] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3982–3992, 2019.

[2] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal Detection of Changepoints With a Linear Computational Cost,” Journal of the American Statistical Association, Vol. 107, No. 500, pp. 1590–1598, 2012.

[3] 堺雄之介, 伊東栄典, “動画サイトにおける視聴者コメントの特徴抽出,” JSAI Technical Report, SIG-KBS, Vol. 124, pp. 17–22, 2021.

[4] “Unsupervised Text Segmentation via Kernel Change-Point Detection on Sentence Embeddings,” arXiv preprint arXiv:2601.18788, 2025.

[5] “Sentiment analysis for live video comments with variational residual representations,” Computer Speech & Language, Elsevier, 2025.

[6] “Multi-modal Video Topic Segmentation with Dual-Contrastive Domain Adaptation,” Proceedings of International Conference on Multimedia Modeling (MMM), pp. 408–421, 2024.

[7] Sui, T., “Measurement and Sentiment Analysis of YouTube Video Comments,” Master’s Thesis, University of Minnesota, 2022.

[8] C.-F. Hsu, E. Khabiri, and J. Caverlee, “How useful are your comments? Analyzing and predicting YouTube comments and comment ratings,” Proceedings of the 19th International Conference on World Wide Web (WWW), pp. 461–470, 2010.

Yahoo! 知恵袋に投稿された回答で参照された web ページの役割の分析

森高 楓† 松香 敏彦‡
†‡千葉大学 〒263-8522 千葉市稲毛区弥生町 1-33
E-mail: †25dm1115@student.gs.chiba-u.jp

あらまし コミュニティ Q&A サイトである Yahoo!知恵袋では、あるユーザの質問に対して、別のユーザが回答する際に、外部の web ページを参照する URL を回答文に含めることがある。参照先の web ページは、回答の根拠や具体例を示すことが多いとされているが、回答文に web ページの内容がどの程度反映されているかは確かめられておらず、回答において参照先の web ページが果たしている役割の詳細は明らかでない。そこで、本研究では、回答文を構成する語が参照先の web ページに含まれるか否かに注目して分析した。回答文において，参照先の web ページに含まれる語とそうでない語が混合されて用いられていたことから，回答文で参照される web ページが，質問の解決に必要な要素を一部提供している可能性が示唆された。

キーワード ユーザ理解，情報利用

1. はじめに

Yahoo!知恵袋のように，ユーザ同士が相互に質問し，回答しあうコミュニティ Q&A サイトに寄せられる投稿文は、人々の情報行動を示すデータとして多くの研究に用いられている。しかし、今までの研究において、コミュニティ Q&A サイトへの投稿文は、ある質問に対する正答例、もしくは質問者の状況を表した説明としてのみ扱われることが多く、コミュニティ外の情報資源の活用行動を示すデータとしては十分に分析されてこなかったように思われる。

そこで、本研究では、「Yahoo!知恵袋」ユーザの情報行動に注目し，質問・回答の過程でコミュニティ外の情報資源がどのように組み込まれ、活用されているかを検討する。特に，回答者が、インターネット上でアクセス可能な web サイトをどのように活用して質の高い回答を作成しているのか明らかにすることは、ユーザの情報探索・利用行動のより深い理解につながると同時に、ユーザの情報ニーズを満たすためのサービスの設計に役立つと考えられる。

2. 関連研究

コミュニティ Q&A サイトにおける情報源の利用について、主にプログラミング技術に特化した Q&A サイトである Stack Overflow を対象とした研究は複数存在する。例えば、Baltes らは、Stack Overflow への投稿に含まれるリンクの参照形態と参照意図を人力でアノテーションして分析し、Stack Overflow でリンクを使用する意図が、紹介したコードスニペットなどの出典を示すことから、主張の根拠を示すことや、使用を進めるツールを紹介することまでのスペクトラムで考えられることや、リンク作成時に参照先ページの内容を引用などで紹介しないものが圧倒的に多いことを指摘した[1]。

一方で、さまざまなトピックが扱われる Yahoo!知恵袋においては、添付される URL についての検討はさほど進んでいない。ただし、投稿された回答のうち，URL を用いて外部の web ページを参照しているものに注目した研究では，URL を用いない回答よりも URL を用いた回答の方がベストアンサーに選ばれた割合が高い一方で，参照するサイトによってベストアンサーへの選ばれやすさが異なることが示された[2]。また，同研究において，Yahoo!知恵袋の回答中で頻繁に参照されるサイトには，おすすめの商品を示すために参照されることが多いサイトと，回答内容の根拠を示すために参照されるサイトの 2 種類があることも指摘されている[2]。しかしながら，個別の回答において，参照された web ページに含まれる記述の引用や転載と，回答者自身の見解や推論が，どのような比率や順番で組み合わせられているかなどの具体的な在り方は，議論の対象となっていない。

3. 本研究の目的

本研究では，先行研究と同様に，Yahoo!知恵袋において web ページを参照する URL を含む回答に注目し，その回答の中で，参照先の web ページに由来する情報と，回答者独自の見解とが，どのように組み合わせられているかを明らかにすることを目的とする。

特に，回答文における web ページの役割を明らかにする端緒として，質問文・回答文・参照先の web ページのそれぞれに含まれる語(形態素)の包含関係に注目することで，簡易的に調査を行うことが本研究の特色である。もし，参照先の web ページに含まれる語が回

答文に含まれないのであれば，参照先の web ページは回答において補助的な役割しか果たしていないと推測できる。一方で，回答文に参照先の web ページに含まれない語が全く存在しないのであれば，参照先の web ページが回答のほぼすべてであり，回答者は適切な web ページを紹介したに過ぎないと推測できる。あるいは，回答文に，参照先の web ページに含まれる語と，そうでない語(つまり，回答者が独自に使用した語)が共に含まれるのであれば，参照先の web ページが回答における根拠や例として組み込まれている可能性が残るだろう。

なお，URL を含む回答投稿であっても高く評価されるとは限らないことを踏まえ，本研究では，分析対象を，質問者を十分に満足させたと思われる回答に限定し，質の高い回答における外部の web ページの利用の実態を検討する。

4. 方法

4.1 対象データ

2018 年 4 月から 2021 年 3 月までの間に Yahoo!知恵袋に投稿された質問とそれに対する回答から 10%をランダムサンプリングして提供されている Yahoo!知恵袋データ（第 3 版） [3] を使用した。

提供データに含まれる 217 万件の質問とそれに対する回答のうち，次の 4 種類の基準のすべてを満たす 1.3 万件の質問と，その質問をしたユーザを満足させたと考えられるかつ URL を含む回答を，実験に使用した。

- **URL 参照**　回答中に URL を 1 つだけ含む。
- **高満足度**　質問者が，回答募集〆切日よりも十分に早い，質問投稿から 6 日未満の段階でベストアンサーに選定した。
- **簡単**　質問と回答がテキストを主体としている，つまり，質問や回答に画像が含まれず，質問には URL も含まれない。また，添付されている URL が動画像を主体とするサイトのドメインではない。
- **分析可能**　回答で参照される URL へのアクセスが可能であり，該当の web ページの本文が形態素解析器で一度に解析可能な長さ以下である。ただし，URL へのアクセスがリダイレクトされ，リダイレクト前後で URL のパスに含まれる英数字の並びが一致する場合は，リダイレクト後の URL を用いた。

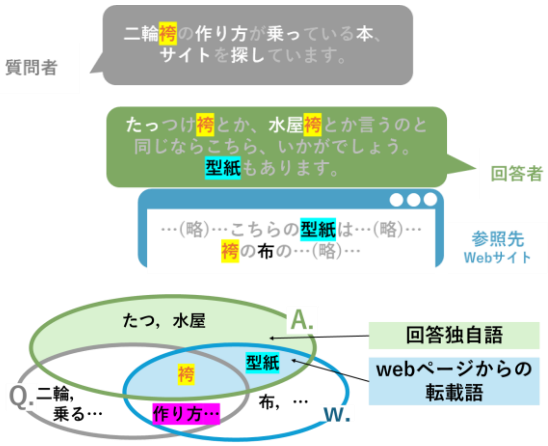
4.2 手続き

まず，分析対象の質問・回答，および回答に含まれる URL から取得した web ページの本文を，テキスト形態素解析機 SudachiPy と解析用辞書 NEologd を用いて形態素解析した。なお，回答に含まれる web ページの本文は，web ページ全体の HTML データを，Python ライブラリの trafilatura を用いて処理することで取得した。

続いて，質問・回答・回答で参照されている web ページの組それぞれについて，含まれる形態素を比較し，回答と回答で参照されている web ページの両方に含まれる語（以下，「web ページからの転載語」とする）と，回答のみに含まれ質問および回答で参照されている web ページには含まれない語（以下，「回答独自語」とする）の種類数を調べた。 (図 1)

なお，この際に考慮する語は，品詞が，名詞・動詞・形容詞のいずれかであるものに限定した。加えて，形態素解析器が持つストップワードリスト・および日本語で良く用いられる Slothlib[3]のストップワードリストのいずれかに含まれる語は分類から除いた。

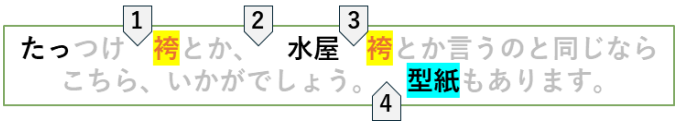
図 1 対象データと形態素の比較分析の例



※ データ例中の灰字は分析対象外の語

加えて，web ページからの転載語・回答独自語の両方を含む回答については，回答の中で，web ページからの転載語が連続する部分と，web ページに含まれない語が連続する部分が何度切り替わるか（以下「転載部・独自部の切替数」）についても調べた。(図 2)

図 2 転載部・独自部の切替数のカウント例
（本例の切替数=4）



※ ハイライトのある語が web ページからの転載語
※ 灰字は分析対象外の語

5. 結果

今回の分析対象の中では，参照している web ページ

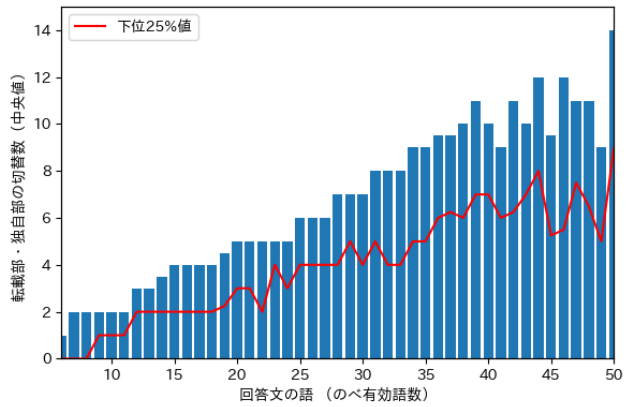
で示された情報を転載しつつ，回答者独自の言葉を付け加えるような回答が最も多かった。しかし，参照している web ページに含まれる情報のみを示す回答や，逆に，参照している web ページに含まれる情報を一切転載しない回答もあった。(表 1)

表 1 web ページからの転載語・回答独自語の有無ごとの事例数

		Web ページからの転載語	
		有	無
回答独自語	有	8834	1950
	無	1027	932

また，web ページからの転載語・回答独自語の両方を含む回答において，回答文が長くなると，転載部・独自部の切替数も増加する傾向が見られた。(図 3)。

図 3　回答文長と転載部・独自部の切替数の関係



6. 考察・結論

今回調査した事例においては，回答文に，参照先の web ページに含まれる語と，そうでない語(つまり，回答者が独自に使用した語)が共に含まれる回答が圧倒的に多いことから，それらの回答において，参照されている web ページは単に直接的な回答を示すものでも，逆に関係性の薄い補足的なものでもなく，質問の解決に必要な一要素としての役割を果たしている可能性が示唆される。加えて，回答文長が長くなると，転載部・独自部の切替数が増加することから，回答文において，web サイトに記載されている情報と回答者独自の情報は複雑に組み合わせられていると考えられる。

以上から，Yahoo!知恵袋に投稿された回答のうち，外部の web ページを参照し，かつ質問者からの評価が高い回答には，質問を解決するために効果的にインターネット情報源を活用している事例が多く含まれる可能性は否定されず，今後さらにその実態を検討する意義があると考えられる。特に，本研究では，語の完全な一致に基づく分析のみ行ったため，文脈を考慮できる分析手法を用いた，より緻密な検討が求められる。

謝　辞

本研究では，国立情報学研究所の IDR データセット提供サービスにより LINE ヤフー株式会社から提供を受けた「Yahoo! 知恵袋データ（第 3 版)」を利用した．

参　考　文　献

[1] S. Baltes, C. Treude and M. P. Robillard, (2020) “Contextual Documentation Referencing on Stack Overflow” IEEE Transactions on Software Engineering, pp. 135-149, 2020.

[2] 杉田飛路，中嶋邦裕，梅本顕嗣，渡辺靖彦，岡田至弘，“Yahoo!知恵袋に投稿された URL を参照している回答の分析” Web インテリジェンスとインタラクション研究会予稿集 2012.

[3] LINE ヤフー株式会社，“Yahoo! 知恵袋データ（第 3 版)”， 国立情報学研究所情報学研究データリポジトリ 2023.

[4] 大島 裕明・中村 聡史・田中克己，“SlothLib Web サーチ研究のためのプロ グラミングライブラリ”，日本データベース学会 letters, 6(1), 113-116 2007

Steam ユーザーの口コミを利用したゲームのプレイ体験の効果的な可視化

相原　颯真†　　横山　昌平†

† 東京都立大学システムデザイン学部情報科学科　〒 191-0065 東京都日野市旭が丘 6-6
E-mail: † aihara-soma@ed.tmu.ac.jp, shohei@tmu.ac.jp,

あらまし　近年，ビデオゲーム市場は拡大を続けている．実際に，PC ゲームの売買プラットフォームである Steam では新作ゲームのリリース数が 2019 年以降右肩上がりとなっている．そのため，ユーザー側から見ると自分の求めるゲームを探すためのハードルが高くなっていることが考えられる．そこで本稿では，ゲームのレビューからユーザーの「プレイ体験」を可視化し，もとから Steam に存在するストアページの検索システムと比較し有用性を示すことでその効果を検証する．

キーワード　レビュー，プレイ体験

1　はじめに

SteamDB[1]によると，世界最大級の PC ゲームを配信するプラットフォームである Steam[2]では，2019 年以降リリースされる，インディーを含めた新作ゲームの数が右肩上がりとなっている [1]．このことからもわかるように，ビデオゲーム産業は，ハードウェアや開発環境の発展，コロナ禍を経た人々の生活スタイルの変化などにより大きな発展を遂げている．

このような市場の拡大は，ユーザーに多様な選択肢を提供する．その一方で，ユーザーが自身の嗜好やプレイスタイルに合致したゲームを効率的に発見することが困難になるという問題も生じさせている．Steam のストアページにはジャンルやタグ，レビュー評価，価格といった項目を用いた検索・絞り込み機能が提供されている．

しかしながら，同一のジャンルであっても没入感，操作感，世界観，難易度などに違いが出るようにユーザーの「プレイ体験」に大きな違いをもたらす可能性があり，公式のストアページから得られる情報はこれらを直接的に反映できるとはいいがたい．

そこで，本研究ではユーザーのプレイ体験によりゲームをフィルタリングする機能及び Steam 上に投稿されたゲームレビューから特徴語を抽出しタグクラウドにする機能を持ったゲームの可視化システムを提案する．また，提案手法を既存の Steam ストアページの検索システムと比較することでユーザーのゲーム検索における有用性を検証する．

公式ストアページにおける「タグ」機能はゲームの開発者及びユーザーがそのゲームについての情報を付け加える機能であり，比較的このプレイ体験に近いものである．本研究では自動でプレイ体験を抽出すること，及びタグクラウドによる情報を評価対象とすることにより差別化を図っている．

以降，第 2 章では本研究に関連する既存研究について述べる．第 3 章では，本研究で提案するシステムの概要および構成について説明する．第 4 章では，提案システムの有効性を検証するために行った実験およびその結果を示す．第 5 章では，実験結果に基づいた考察を行う．第 6 章では，本研究のまとめと今後の課題について述べる．

1：https://steamdb.info/（参照 2025-12-17）
2：Steam, https://store.steampowered.com/,（参照 2025-12-17）

2　関連研究

近年のインターネットの発達から EC サイトをはじめとしたレビューや SNS の投稿といったテキストデータが大量に蓄積されている．そのため，これらの非構造データからからテキストマイニングによってデータの分析や可視化をする試みは数多く存在する．

岑ら [2] は，映画レビューをワードクラウドによって可視化する試みを行っており，コロナ禍前後のレビューを比較することで特徴語の変化をとらえている．ワードクラウドの出現頻度の比較を行うことにより，語彙の変化を視覚的にとらえられるようになったとしている．本研究においても，データ可視化の手法としてタグクラウドを用いた．

北村ら [3] は，コスメアイテムのレビューをタグクラウドによって可視化し，タグクラウドの要素にカーソルを合わせると関連するレビューが見られるアプリケーションを作成した．単にレビューの文章を並べるのに比べタグクラウドの方がユーザーの意思決定にかかる時間が短く，ストレスを感じにくいという実験結果となった．好意的なレビューと否定的なレビューのタグクラウドを並べる点，タグクラウドに現れる単語に関連するレビューを表示するという手法を本研究でも取り入れた．

小川 [4] は，映画のレビューをワードクラウドと共起ネットワークを用いて可視化をした．ワードクラウド上に出現した単語どうしのつながりを共起ネットワークを用いて抽出することにより，レビューデータという非構造的なデータを映画評価に関連するものに集約させ，これらの分析をより意味のあるものとしている．評価の高い作品に対し可視化を行うことにより，ヒット作に共通した要素を抽出できたと結論付けている．

プレイ体験を抽出するにあたり，アスペクト抽出という分野を参考にした．アスペクト抽出は特定のデータからある特徴や

側面を抽出するためのプロセスである．レビューにおいて，全体の感情だけでなく，アスペクトごとの感情を用いることでよりレビューにおける感情分析を正確に行うことが主な目標とされている．レビューという構造化されたデータから「音楽」「難易度」といった側面を抽出し，新たに分析対象とするものが，本研究におけるアスペクト分析である．これに関連した研究を以下に示す．

Bin ら [5] は，教師あり学習で使用されるアルゴリズムである線形判別分析（LDA）を拡張し，各単語はアスペクトと感情によって生成されるとした．これらを自動でタグ付けし，アスペクト抽出と感情分類を同時に行うモデルを提案することを目標とした．通常の LDA と比較して精度が高く出たほか，典型的なアスペクトに関しては特に安定して抽出できたとしている．

Najwa ら [6] は，映画，レストラン，飛行機という 3 つの異なる分野でのレビューにおいて，特定の観点を表しているような単語群，つまりアスペクトを自動で抽出し，それを利用した感情分類モデルを作成した．事前知識を取り入れた LDA である SA-LDA，感情辞書 SentiWordNet を用いた単語への自動ラベリング，アンサンブル分類器を用いた．従来のモデルよりも性能が向上したと述べられている．本研究では，アスペクトラベルの付与を手動で行ったものの，単語の感情スコアをタグ付与に用いるという点を参考にした．

これらの研究は，レビューのような非構造データをテキストマイニングし，特徴語を抽出することによる分析の有用性を示している．加えて，タグクラウドや共起ネットワークといった可視化手法を用いることによって，分析によって得た情報をわかりやすく可視化できていることを示せている．こういった研究は，ゲーム以外のレビューに焦点を当て，情報の可視化を行ったものとなっているが，ゲームのレビューについての研究も存在する．以下に具体的な関連研究を示す．

Tibor ら [7] は，英語データの Steam レビューに関する分析を行っている．スクレイピングによってレビューを取得し，感情分析を行ったのち語数，ジャンル，レビュー内での感情変化などの軸から分析を行った．プレイ時間が長いほどポジティブなレビューが書かれやすかったり，ジャンルごとに特徴的な感情が傾向があることなどが示唆されている．

廣田 [8] は，Steam 上のゲームレビューに対し，ジャンルごとに感情分析を行っている．Plutchik の感情の輪より 8 つの感情に各レビューを分類することでジャンルごとにプレイヤーが持ちがちな感情の傾向を明らかにしており，各ジャンルでの感情値の傾向が異なると結論付けられている．

また，廣田ら [9] は，Steam におけるジャンルごとのレビュー観点の分析を行っている．これは，Steam ストアにあるユーザーや開発者によって付与されたタグではなく，独自に「音楽」「ストーリー」といった観点を設定し，LLM を用いて各レビューがこれらの要素をどのように含むかを列挙するという手法を提案した．ゲームのジャンルによって言及されるレビュー観点が異なったり，逆にジャンル間に共通するレビュー観点があることを発見した．本研究では，このレビュー観点からプレイ体験のアイデアを得た．

以上のように，レビューの分析やタグクラウドを用いた可視化，ゲームのレビューをテーマにした研究は以前より行われている．しかし，これらはそれぞれ独立した研究として存在している．本研究では，これらを融合し，プレイ体験とタグクラウドをもとにゲームを可視化するアプリケーションを作成した．

3 提案システム

本章では，提案システムについて述べる．本システムは，Steam に投稿されたゲームレビューを入力とし，テキスト処理および特徴語抽出を行うことで，ユーザーのプレイ体験を可視化することを目的としている．ここで，本論文では，プレイ体験を，ゲームプレイを通じてユーザーが主観的に感じ取った印象や評価が，Steam レビューにおいてテキストとして表現された内容の集合として定義する．Steam ストアにおけるジャンルは,「RPG」や「アクション」など主にゲームの構造的特徴に基づく分類であるのに対し，本研究で扱うプレイ体験は,「音楽に特徴がある」「チームワークが重要」などユーザー視点の体感的特徴を反映することを目的とした．

さらに，感情分析によってポジティブな文脈で使われているもののみを抽出した．これは，プレイ体験によるフィルターをポジティブな方向のみに機能させるためのものだからである．ユーザーがゲームを選ぶ際に音楽を重視するとしたとき，わざわざ音楽にネガティブなレビューを多く持つゲームを提示する必要はない．あくまで，ユーザーが加点方式でゲームの選択を行うことを想定しアプリケーションの作成を行ったため，このような設計とした．図 1 に，提案システムの全体構成を示す．以下 3.1 節から 3.6 節において図 1 のフローチャートにおける各項目を説明する．

3.1 SteamAPI によるゲームデータ及びレビューの取得

まず，分析対象としてプレイヤー数が多いゲームを抽出するために SteamCharts[3]から実装時点（2025/10/21/19:25）でプレイヤー数が上位 1000 件だったゲームの名前，appid（ゲームを識別するための一意性を持った ID）を取得し，JSON ファイルに保存した．しかし，実行中のエラーにより実際に取得できたゲームは 969 件となっている．続いて，この JSON ファイルに保存された appid を入力として Steam Web API[4]から必要となるデータを取得した．具体的にはゲームのレビュー，ジャンル，値段が無料であるか，無料でないならその値段である．レビューの取得は，その後のデータ分析の行いやすさ及び Steam のユーザー人口を考慮し英語で行った．

SteamAPI では，レビュー取得の際，ユーザーが投稿したレビューに対し，他のユーザーが参考になったと投票した数が多い順にソートした「参考になった順」を使用することができる．また，ユーザーはレビューを投稿する際にそのゲームを「おすすめする」「おすすめしない」どちらかを選択する．本研究では，

3：https://steamcharts.com/,（参照 2025-10-21）.

4：https://partner.steamgames.com/doc/store/getreviews,（参照 2026-1-2）.

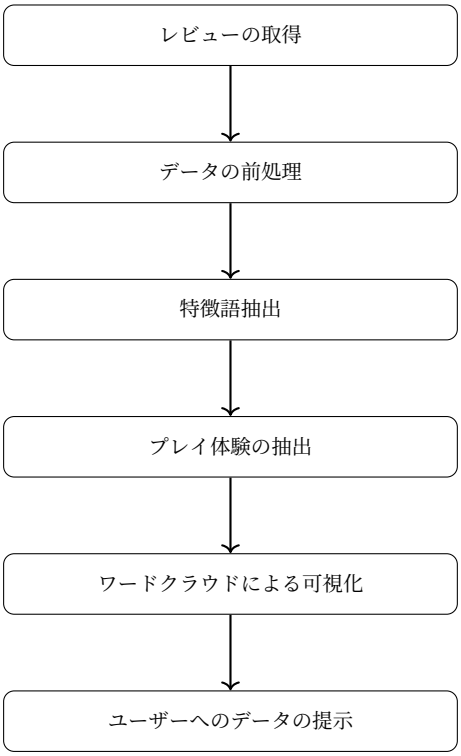


図 1　本研究における処理の流れ

各ゲームに対し，参考になった順にソートしたものを「おすすめする」「おすすめしない」それぞれ上位 100 件ずつを取得した．ゲームによってレビュー数がバラバラであるためレビュー数によって評価にブレを生じさせたくないと考えたため，ゲームごとのレビュー数は 100 件という比較的少ないデータ量である．レビューを参考になった順にソートしたため各レビューにある程度の説得力が確保できると考えた．しかし，レビュー数の都合上 100 件のレビューデータが取れなかったものもあった．加えて，各ゲームにおけるヘッダー画像も取得した．

3.2　データの前処理

続いて，データの分析を行いやすくするためにデータの前処理を施した．まず，表記ゆれを防ぐため全単語を小文字に変換し，同一語が別単語として検出されるのを防いだ．また，単語による分析には必要ない絵文字のような特殊文字や数字を除外した．加えて，形態素解析を行った．本研究で扱うレビューは英語であるため，形態素解析は単語分割のみを行った．

最後に，特徴語抽出を行う際にノイズとなりがちな普遍的な単語及びゲームの略称となるものをストップワードとして除外した．普遍的な単語として除外したストップワードのリストを表 1 に示す．ゲームの名称に関しては,「そのゲームに含まれる単語，およびその単語同士を組み合わせたもの」として取得した．例えば,「Monster Hunter Wilds」というゲームでは,「monster」「hunter」「wilds」及びその組み合わせから得られる「monster hunter」「hunter wilds」「monster hunter wilds」という単語を除外語として設定した．

あるゲームのレビュー内でそのゲーム名の名前を含む文章が書かれることは珍しくない一方，その理由で出てきた単語は他

表 1　ストップワード一覧

ストップワード
game, play, played, playing, player, players, best, really, very, lot, much, many, hours, hour, long, short, thing, stuff, everything, good, bad, great, recommend, recommended, recommendable, buy, buying, games, one, still

文書には出現しにくい．そのため，全文書と比較した TF-IDF 分析を行った場合ゲーム名に由来する単語の TF-IDF スコアが大きく出てしまう可能性がある．このような単語はゲーム内容に対する意味を見出しづらいため除外をしたい．しかしながら，他のゲームを対象としたレビューででこれらの単語が出てきた場合，そのゲームと比較していたり，類似性がある可能性が考えられる．したがって，本研究では，あくまで各ゲームについて分析を行う場合のみそのゲームに登場する単語を除外することとした．

3.3　特徴語抽出

本研究では，特徴語の抽出を行うために，TF-IDF 法を用いた．TF-IDF 法はある文書内における単語の重要性を求めるために役立つ指標の一つであり，ある単語がある文書の中でどれほど重要かを表す Term frequency（単語頻度，以下 TF）と，ある単語が文書全体でどれほど重要かを表す Inverse document frequency（逆文書頻度，以下 IDF）を掛け合わせることによって表される．$n(t,d)$ を文書 d 中に含まれる単語 t の出現回数とし，K を全単語集合，$\sum_k^K n(k,d)$ を文書 d 中に出現する全単語の出現回数の和，$|D|$を文書の総数，$df(t)$ を単語 t が現れる d の数，TF(t,d) を単語 t の文書 d における出現頻度，IDF(t) を特定のドキュメントのみに出現する傾向の高さとすると，文書 d 中に含まれる単語 t の重要度 TFIDF(t,d) は，以下のように求められる．

$$TF(t,d) = \frac{n(t,d)}{\sum_k^K n(k,d)} \tag{1}$$

$$IDF(t) = log(\frac{|D|}{df(t)}) \tag{2}$$

$$TFIDF(t,d) = TF(t,d) * IDF(t) \tag{3}$$

前処理を行った各ゲームのレビューに対し TF-IDF 法を用い，全レビューと比較して TF-IDF スコアが高い特徴的な単語を抽出し、おすすめする・おすすめしないのレビューごとに上位 10 単語を記録する．つまり，式 (1) から (3) において，文書 d を目的ゲームのレビュー及び他ゲームのレビュー，単語 t を目的ゲームのレビューに含まれるすべての単語としている．

特徴語抽出は n=1,2 の n-gram として行った．例えば「BGM」という単語があったとき，そこから得られる情報は「BGM に特徴がある」というものしかない．しかし，もし「chill BGM」として抽出されると,「落ち着いた BGM が使われている」というところまで得られる情報が増える．つまり，それと関連する語，つまりレビュアーがどのような意図でその単語を使ったのかについて抜き出しやすくするため，n=1,2 の n-gram を採用した．また，抽出対象となる単語を名詞と動詞のみに絞ること

で，具体的な語を効率的に得られると考えた．

3.4 プレイ体験の抽出

プレイ体験のタグ（以下タグ）の設定は，手動で行った．タグは，3.1 節の前提を基に，[9] のレビュー観点を参考に作成した．タグの一覧を表 2 の左辺に，各ゲームにタグを割り当てるアルゴリズムのフローチャートを図 2 に示す．また，タグ付けを行うためにタグと関連するような単語をタグ関連語として設定する．タグ関連語の一覧を表 2 の右辺に示す．タグを各ゲームのレビューから抽出するために，まず各ゲームのレビューに対し各タグに対する関連単語が出現する回数及びそこで出現した単語の感情スコアより平均感情スコアを求める．感情スコアは感情分析によって計算する．ここから「単語の出現率」×「平均感情スコア」を計算する．

この値を各ゲーム各単語の単語スコアと定義し，Percentile ベースによる評価を行う．まず，各単語スコアが全体のゲーム 969 件中何位に相当するかを計算し，上位何割かに当たる数字（Percentile）を求める．そして，閾値によるタグ付けとして，閾値及び Percentile が一定以上のタグを各ゲーム順に割り当てる．それぞれが一定以上の値であることにより，タグがそのゲームに関係あることが保証される．この操作を閾値と Percentile を小さくしてもう一度行う．これでもタグが付かなかったゲームがある場合，単語スコアが最も高いタグを付与することにより，際立った特徴がないゲームに対してもタグの付与ができる．最後に，タグの上限数によるカットを行う．タグの上限を決め，それより多いタグが付いてしまった場合下位のものを削除する．特定のゲームに多すぎるタグが付いてしまった場合，どのタグで検索してもそのゲームが表示されることを防ぐためである．

図 3 にジャンルを縦軸，付与されたタグの割合を横軸としたヒートマップを示す．Content Volume や Game Mechanics，Player Skill は多くのジャンルにおいて現れていることからこれらの要素はどんなゲームを開発する際に重要であり，story は Action や Adventure，RPG において現れていることから，これらのジャンルのゲームを開発する際に意識すべきということがわかる．

3.5 タグクラウドによる可視化

前項のようにして求めた特徴語を，タグクラウドを用いて可視化した．タグクラウドは文書集合中に出現する語をあらかじめ設定したパラメータの大きさに応じて視覚的に表現するものである．パラメータの大きい項目ほど大きい文字サイズで表示されるため，データに含まれる特徴を直感的に理解することができる．本研究では，TF-IDF スコアをパラメータとしてタグクラウドを作成した．したがって，そのゲームにおいて特徴的な語がタグクラウドの中で大きく出力される．

タグクラウドは，3.1 節において取得したレビューの「おすすめする」「おすすめしない」それぞれについて作成した．つまり，各ゲームに 2 つずつのタグクラウドが作成される．これにより，ユーザーはそのゲームのよい点，悪い点をゲームの各

表 2 タグ一覧

タグ	関連語
Game Mechanics	mechanic, mechanics, gameplay, control, combat, system, strategy
Game Balance	balance, balanced, overpowered, op, nerf, buff, fair, unbalanced
Music	music, soundtrack, sound, audio, score, composer, theme, bgm
Story	story, narrative, plot, character, dialogue, lore, campaign, story-driven
Immersion	immersion, immersive, atmospheric, atmosphere, immerse, engaging, absorbing
User Interface	ui, interface, menu, button, navigation, intuitive, user interface
Graphics	graphics, visual, art, animation, model, texture, beautiful, gorgeous, visual design
Community	community, multiplayer, social, cooperative, online, players, friendly
DLC	dlc, downloadable content, expansion, dlcs, bonus content, extra content
Mods	mod, mods, modding, modifiable, customization, moddable
Content Volume	content, hours, playtime, replay, endless, roguelike, length, amount
Player Skill	skill, difficulty, challenge, hard, easy, learning curve, require skill, skill-based

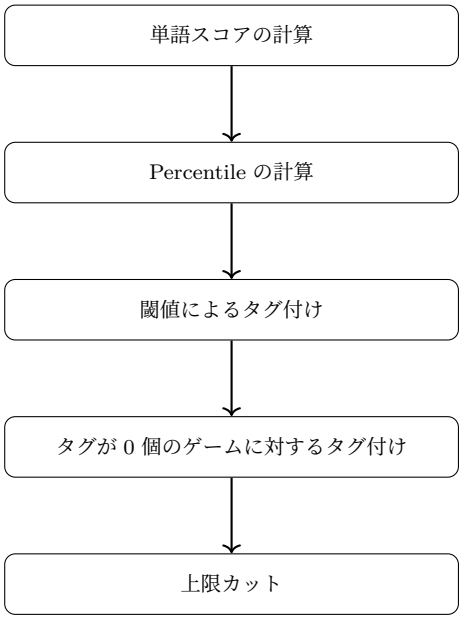


図 2 本研究における処理の流れ

タグクラウドを見ることにより判断することができる．タグクラウドは Python の wordcloud ライブラリ [5]によって作成した．作成したタグクラウドの例を図 4 及び図 5 に示す．これは，

5：https://pypi.org/project/wordcloud/,（参照 2025-12-22）.

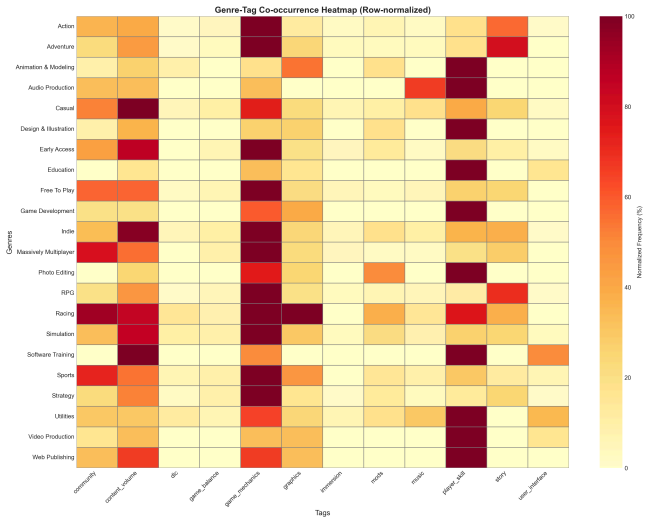


図 3　付与されたタグとジャンルのヒートマップ

図 4 「おすすめする」のタグクラウド

PUBG: BATTLEGROUNDS[6]について作成したタグクラウドであり，図 4 が「おすすめする」，図 5 が「おすすめしない」のレビューについて分析を行った結果である．

これらのタグクラウドから，例えば図 4 には「battle royale」「friends」という単語が含まれているため，このゲームは友人と遊ぶことができ，バトルロワイヤル形式のゲームであることがわかる．また，図 5 には「cheaters」という単語が含まれていることから，ゲーム内にチートを使用する人間が存在することがわかる．このように，タグクラウドを一目見ただけでそのゲームの特徴を把握することを目標とした．

3.6　ユーザーへのデータの提示

前項で作成したタグクラウド及び 3.1 節において取得したヘッダー画像を用い，ジャンル，価格，及びプレイ体験によってゲームの検索ができるアプリケーションを作成した．フレームワークには React を採用し，JavaScript を用いて実装を行った．以下に，各画面の持つ機能について説明を行う．

まずアプリケーションに接続するとトップ画面が表示され

6：https://store.steampowered.com/app/578080/PUBG_BATTLEGROUNDS/,(参照 2026-1-24)

図 5 「おすすめしない」のタグクラウド

る．この画面はフィルター及びソートを行い，目的となるゲームを探すことを目的としている．「フィルター・ソート」の部分でゲームの金額・ジャンルによるフィルター，指定範囲内の金額の中で昇順，降順のソートを行うことができる．デフォルトでは SteamCharts からデータを取得した際の同時接続プレイヤーが多かった順にソートされている．このようにしてフィルター・ソートされたゲームがゲーム名，ジャンル，値段，ヘッダーとともに表示され，ユーザーは提示されたゲームの中から目的のものを探す．本アプリケーションは特定のゲームを探すことを目的としておらず，あくまでユーザーが求めるプレイ体験からそれに合うものを探すという目的で作られているため，ゲーム名による検索機能など Steam 公式ストアに存在する検索機能の一部を省いている．

この中のいずれかのゲームをクリックするとゲーム詳細画面へと遷移する．ゲーム詳細画面では，ユーザーが選択したゲームのタグクラウド及び関連するレビューが閲覧できる．各ゲームについて，3.5 節において作成した「おすすめする（Recommended）」「おすすめしない（Not Recommended）」の 2 種類のタグクラウドが表示される．加えて，下部分に取得したレビューが表示される．このレビューはタグクラウドに存在する単語を指定し，その単語を含むレビューのみをフィルターできるようになっている．ユーザーはこの 2 種類のタグクラウドに含まれる単語及びレビューから，そのゲームに対する評価を行う．

4　実　　験

提案したアプリケーションの有用性を評価するため，東京都立大学の大学生と大学院生合わせて 5 名，筆者の友人である大学生 5 名の合計 10 名を対象に評価実験を実施した．以下に実験の概要及び実験結果を示す．

被験者は，提案手法（本研究で作成したアプリケーション）と既存手法（Steam 公式ストア）を用い，13 個の質問に回答する．Q2 及び Q4 にあるゲームを探すのにかかった時間を計測するタスクでは，各リンクをクリックしてから目的となるゲームを探し終えるまでの時間を計測した．被験者を A 班 5 名と

B 班 5 名に分け，提案手法と既存手法で異なるゲームを探すことにより結果の偏りを防いだ．質問の内容を以下に示す．
Q1: あなたは普段週に何時間くらいデジタルゲーム（スマホ、PC、タブレットなど）をプレイしますか？
Q2: 以下の条件を満たすようなゲームを A 班：提案手法 B 班：既存手法で 2 つ探して以下に記入してください。また、検索にかかった時間も測定し記入してください。
・ジャンルが「アクション（Action）である」
・価格が 3000 円以下である
・ゲームバランスがとれている
Q3:Q2 で見つけたそれぞれのゲームについて、レビューやタグクラウドを見てよい点、悪い点を探し、挙げてください。（よい点：マルチプレイが楽しい、悪い点：チーターが多いなど）
Q4: 以下の条件を満たすようなゲームを A 班：既存手法 B 班：提案手法で 2 つ探して以下に記入してください。また、検索にかかった時間も測定し記入してください。
・ジャンルが「RPG である」
・価格が 2000 円以下である
・ストーリーに特徴がある
Q5:Q4 で見つけたそれぞれのゲームについて、レビューやタグ、概要を見てよい点、悪い点を探し、挙げてください。（よい点：マルチプレイが楽しい、悪い点：チーターが多いなど）
Q6: プレイ体験によるフィルターは希望のゲームを見つけるのに役立ちましたか？
Q7: ワードクラウドとレビュー検索機能はゲームの特徴を理解するのに役立ちましたか？
Q8: 提案手法は、複数のゲームの違いを視覚的に比較しやすいと感じましたか？
Q9: どちらのほうが詳細情報に素早くアクセスできると感じましたか？
Q10: ゲーム情報を視覚的にわかりやすく表示できていたのは、どちらでしたか？
Q11: どちらのほうが使いやすいと思いましたか？
Q12: 提案手法の UI についての感想をよい点/悪い点どちらでも構わないので教えてください。
Q13: 実験に関して不明点・感想等があれば教えてください。

Q1 では被験者がどれくらいゲームをする習慣があるのかを問う．選択肢は「週 7 時間未満」「週 7 時間以上 21 時間未満」「週 21 時間以上」の三択とした．. 結果は,「週 7 時間未満」が 3 名,「週 7 時間以上 21 時間未満」が 2 名,「週 21 時間以上」が 5 名であった．

Q2 から Q5 では被験者に実際に提案システムと既存手法を使用しゲームを探させた．検索に要した時間を測り，見つけたゲームに対してよい点，悪い点を探させることによりゲームの特徴が伝わっているのかを確かめた．全被験者について，検索にかかった時間は提案手法の方が短かった．検索に要した時間に被験者間で大きなブレがあるため平均値ではなく中央値で比較すると，提案手法が 72 秒に対し既存手法は 117.5 秒であった．図 3 に被験者ごとにタスク完了までにかかった時間をまとめた表を示す．

表 3　被験者ごとの検索時間（秒）

被験者	提案手法	既存手法
1	33	47
2	81	85
3	330	420
4	60	473
5	63	66
6	30	43
7	30	50
8	129	271
9	193	348
10	105	150
中央値	72	117.5

Q6 から Q8 は提案手法の有用性を主観的に問う設問となっている．各設問は 5 件法リッカート尺度を用いて設計した．Q6 および Q7 では有用性について,「全く役立たなかった」を 1 点,「非常に役立った」を 5 点とする 5 段階評価とした．Q8 では主観的印象について,「全く感じなかった」を 1 点,「非常に感じた」を 5 点とする 5 段階評価とした．回答結果は数値化したうえで平均値を算出し，設問ごとの評価指標とした．

Q6 では「普通」が 3 人,「役立った」が 4 人,「非常に役立った」が 3 人であり，平均値は 4.0 であった．Q7 では「役立たなかった」が 3 人,「普通」が 2 人,「役立った」が 2 人,「非常に役立った」が 3 人であり，平均値は 3.5 であった．Q8 では「感じなかった」が 2 人,「普通」が 3 人,「感じた」が 3 人,「非常に感じた」が 2 人であり，平均値は 3.5 であった．

Q9 から Q11 は提案手法と既存手法を比較し，どちらの方が有用と感じたかを主観的に問う設問となっている．選択肢は「提案手法」「既存手法」とし，そう判断した根拠も回答させた．Q9 では「提案手法」が 6 人、「既存手法」が 4 人，Q10 では「提案手法」が 5 人、「既存手法」が 5 人，Q11 では「提案手法」が 6 人、「既存手法」が 4 人となった．

Q12 及び Q13 は提案手法についての意見を問う自由記述の質問であり，これを今後の展望とすることを意図として設定した．「検索方法や内容が分かりやすい」「タグクラウドを使用することにより一目でレビューの傾向が把握できる」といったポジティブな意見が得られた一方,「ブラウザの戻るボタンで前の画面に戻るとフィルターが初期化されてしまう」「タグクラウドに対する説明がないため有効活用した検索ができない」というネガティブな意見も得られた．

5 考　　察

Q2 から Q5 において，すべての被験者が既存手法と比較して提案手法の方が検索にかかる時間が短くなっており,「ジャンル」「価格」「プレイ体験」という条件のもとゲームを探す場合，提案手法の方が効率的な検索が行えることが分かった．これは，提案手法が要求される指標をすべて盛り込んだ検索を行える一方，既存手法ではプレイ体験という検索項目自体が用意されて

おらず，レビューから判断しなければならず，その分時間にロスが生まれたためであると考えられる．

また，Q6 から Q8 のすべての設問において，平均値が 3 を上回っており，提案手法において実装した「プレイ体験によるフィルター」「タグクラウド及びレビュー検索機能」「複数ゲームの比較」については一定の有用性が示せたといえる．

Q9 から Q11 においては，すべての設問において提案手法と既存手法の評価はほぼ同程度であった．提案手法では最低限の機能のみを実装したが，既存手法では写真や動画が充実しているため視覚的な情報が多く，判断情報になりやすいという意見が得られた．

6 おわりに

本研究では，ゲームレビューに着目し，テキスト分析に基づくプレイ体験の抽出及び可視化を目的とした．Steam より取得したゲームについて TF-IDF 分析などの手法によって特徴語抽出を行い，得られた結果をもとにタグクラウドを用いた可視化を行った．作成したタグクラウド及びレビューからユーザーのゲーム探しに役立つようなアプリケーションを開発し，Steam ストアと比較してその有用性を示した．評価実験によりプレイ体験によるフィルターやタグクラウドによる特徴語の可視化には一定の効果があったといえる．

今後の課題としていくつかの点が挙げられる．まず，レビューデータを英語で取得した点である．Steam コミュニティにおける人口の多さ，テキストマイニングのしやすさから英語を用いて研究を行う判断をしたが，日本語圏と英語圏だと好まれるゲームのタイプが異なる．そのため，日本語圏のユーザーを対象とした実験では適したデータが得られなかった可能性が考えられる．

また，特徴語抽出のアルゴリズム，UI にも改善の余地があると考えられる．TF-IDF 法は文書集合内における単語の出現頻度，希少性に焦点を当てたものであり，文書における特徴的な単語を抽出することに適している．加えて，教師なしで使用可能な点，大規模な学習データがなくてもデータ分析を行うことができる点から，本研究では TF-IDF 法を採用した．

しかしながら，多くのゲームのタグクラウドにおいて「like」が出てきてしまうなど，明確に特徴語の抽出ができておらず，タグクラウドの中でも具体的にゲームの印象をつかめる語自体が限られてしまっており，タグクラウドが真に有効とはいえなかった．ストップワードを調整し，明らかに重複する語を取り除くことによっても語の具体性を高めることができる可能性はあるが，設定者の主観が入ってしまう可能性を鑑みるとアルゴリズムそのものの見直しも重要であると考えられる．

さらに，同時接続人数を基にデータの取得をしたため，分析対象が現在流行中のゲームであったり対戦系のゲームになってしまったことから，取得できるデータにムラができてしまったとも考えられる．

文　献

[1] SteamDB. Steamhttps://steamdb.info/stats/releases/.（参照 2025-12-17）.

[2] 岑天霞 and 渡邉英徳. ワードクラウドの時系列による映画レビューの可視化 感染症をテーマとした映画 『コンテイジョン』を例に. **デジタルアーカイブ学会誌** *2021*, 5(s1):s47–s50, 2021.

[3] 北村 寧来, 横山真由子, 笹原 彰斗, 上田真由美, and 中島 伸介. コスメレビューの特徴量可視化手法の提案. *DEIM2025*, 2025.

[4] 小川 哲司. テキストマイニングとネットワーク分析を用いた映画評価の要因分析. **経済経営論集**, 29(2):26–35, 2021.

[5] Bin Lu, Myle Ott, Claire Cardie, and Benjamin Tsou. Multi-aspect sentiment analysis with topic models. *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 81–88, 2011.

[6] Najwa AlGhamdi, Shaheen Khatoon, and Majed Alshamari. Multi-aspect oriented sentiment classification: Prior knowledge topic modelling and ensemble learning classifier approach. *Applied Sciences*, 12(8), 2022.

[7] Tibor Guzsvinecz and Judit Szűcs. Length and sentiment analysis of reviews about top-level video game genres on the steam platform. *Computers in Human Behavior*, 149(107995):1–14, 2023.

[8] 廣田雅春. Steam におけるビデオゲームの日本語レビューのジャンルごとの感情分析. **日本デジタルゲーム学会**, 15(予稿):145–148, 2025.

[9] 廣田 雅春 and 下田 紀之. Steam におけるビデオゲームの日本語レビューを用いた ジャンルごとのレビュー観点に関する分析. **日本デジタルゲーム学会**, 2025 年 夏季研究発表大会 (予稿):9–14, 2025.